

On finitely generated profinite groups, I: strong completeness and uniform bounds

Nikolay Nikolov* and Dan Segal

February 2, 2008

Abstract

We prove that in every finitely generated profinite group, every subgroup of finite index is open; this implies that the topology on such groups is determined by the algebraic structure. This is deduced from the main result about finite groups: let w be a ‘locally finite’ group word and $d \in \mathbb{N}$. Then there exists $f = f(w, d)$ such that in every d -generator finite group G , every element of the verbal subgroup $w(G)$ is equal to a product of f w -values.

An analogous theorem is proved for commutators; this implies that in every finitely generated profinite group, each term of the lower central series is closed.

The proofs rely on some properties of the finite simple groups, to be established in Part II.

Contents

- §1. Introduction
- §2. The Key Theorem
- §3. Variations on a theme
- §4. Proof of the Key Theorem
- §5. The first inequality: lifting generators
- §6. Exterior squares and quadratic maps
- §7. The second inequality, soluble case
- §8. Word combinatorics
- §9. Equations in semisimple groups, 1: the second inequality
- §10. Equations in semisimple groups, 2: powers
- §11. Equations in semisimple groups, 3: twisted commutators

*Work done while the first author held a Golda-Meir Fellowship at the Hebrew University of Jerusalem

1 Introduction

A profinite group G is the inverse limit of some inverse system of finite groups. Thus it is a compact, totally disconnected topological group; properties of the original system of finite groups are reflected in properties of the topological group G . An algebraist may ask: does this remain true if one forgets the topology? Now a base for the neighbourhoods of 1 in G is given by the family of all open subgroups of G , and each such subgroup has finite index; so if *all* subgroups of finite index were open we could reconstruct the topology by taking these as a base for the neighbourhoods of 1.

Following [RZ] we say that G is *strongly complete* if it satisfies any of the following conditions, which are easily seen to be equivalent:

- (a) every subgroup of finite index in G is open,
- (b) G is equal to its own profinite completion,
- (c) every group homomorphism from G to any profinite group is continuous.

This seems *a priori* an unlikely property for a profinite group, and it is easy to find counterexamples. Indeed, any countably based but not finitely generated pro- p group will have $2^{2^{\aleph_0}}$ subgroups of index p but only countably many open subgroups; more general examples are given in [RZ], §4.2, and some examples of a different kind will be indicated below. Around 30 years ago, however, J.-P. Serre showed that every *finitely generated* pro- p group is strongly complete. We generalize this to

Theorem 1.1 *Every finitely generated profinite group is strongly complete.*

(Here, ‘finitely generated’ is meant in the topological sense.) This answers Question 7.37 of the 1980 Kurovka Notebook [K], restated as Open Question 4.2.14 in [RZ]. It implies that the topology of a finitely generated profinite group is completely determined by its underlying abstract group structure, and that the category of finitely generated profinite groups is a full subcategory of the category of (abstract) groups.

The theorem is a consequence of our major result. This concerns *finite* groups having a bounded number of generators, and the values taken by certain *group words*. Let us say that a group word w is *d-locally finite* if every d -generator (abstract) group H satisfying $w(H) = 1$ is finite (in other words, if w defines a variety of groups all of whose d -generator groups are finite).

Theorem 1.2 *Let d be a natural number, and let w be a group word. Suppose either that w is d -locally finite or that w is a simple commutator. Then there exists $f = f(w, d)$ such that: in any finite d -generator group G , every product of w -values in G is equal to a product of f w -values.*

Here, by ‘simple commutator’ we mean one of the words

$$\begin{aligned} [x_1, x_2] &= x_1^{-1} x_2^{-1} x_1 x_2, \\ [x_1, \dots, x_n] &= [[x_1, \dots, x_{n-1}], x_n] \quad (n > 2), \end{aligned}$$

and a *w-value* means an element of the form $w(g_1, g_2, \dots)^{\pm 1}$ with the $g_j \in G$.

Profinite results

The proof of Serre’s theorem sketched in §4.2 of [Sr] proceeds by showing that if G is a finitely generated pro- p group then the subgroup $G^p[G, G]$, generated (algebraically) by all p th powers and commutators, is open in G . To state an appropriate generalization, consider a group word $w = w(x_1, \dots, x_k)$. For any group H the corresponding *verbal subgroup* is

$$w(H) = \langle w(h_1, \dots, h_k) \mid h_1, \dots, h_k \in H \rangle,$$

the subgroup generated (*algebraically*, whether or not H is a topological group) by all w -values in H . We prove

Theorem 1.3 *Let w be a d -locally finite group word and let G be a d -generator profinite group. Then the verbal subgroup $w(G)$ is open in G .*

To deduce Theorem 1.1, let G be a d -generator profinite group and K a subgroup of finite index in G . Then K contains a normal subgroup N of G with G/N finite. Now let F be the free group on free generators x_1, \dots, x_d and let

$$D = \bigcap_{\theta \in \Theta} \ker \theta$$

where Θ is the (finite) set of all homomorphisms $F \rightarrow G/N$. Then D has finite index in F and is therefore finitely generated: say

$$D = \langle w_1(x_1, \dots, x_d), \dots, w_m(x_1, \dots, x_d) \rangle.$$

It follows from the definition of D that $w_i(\mathbf{u}) \in D$ for each i and any $\mathbf{u} \in F^{(d)}$; so putting

$$w(\mathbf{y}_1, \dots, \mathbf{y}_m) = w_1(\mathbf{y}_1) \dots w_m(\mathbf{y}_m)$$

where $\mathbf{y}_1, \dots, \mathbf{y}_m$ are disjoint d -tuples of variables we have $w(F) = D$. This implies that the word w is d -locally finite; and as $w_i(\mathbf{g}) \in N$ for each i and any $\mathbf{g} \in G^{(d)}$ we also have $w(G) \leq N$. Theorem 1.3 now shows that $w(G)$ is an open subgroup of G , and as $K \geq N \geq w(G)$ it follows that K is open.

The statement of Theorem 1.3 is really the concatenation of two facts: the deep result that $w(G)$ is *closed* in G , and the triviality that this entails $w(G)$ being open. To get the latter out of the way, say G is generated (topologically) by d elements, and let $\mu(d, w)$ denote the order of the finite group $F_d/w(F_d)$ where F_d is the free group of rank d . Now suppose that $w(G)$ is closed. Then

$w(G) = \bigcap \mathcal{N}$ where \mathcal{N} is the set of all open normal subgroups of G that contain $w(G)$. For each $N \in \mathcal{N}$ the finite group G/N is an epimorphic image of $F_d/w(F_d)$, hence has order at most $\mu(d, w)$; it follows that \mathcal{N} is finite and hence that $w(G)$ is open.

Though not necessarily relevant to Theorem 1.1, the nature of other verbal subgroups may also be of interest. Using a variation of the same method, we shall prove

Theorem 1.4 *Let G be a finitely generated profinite group and H a closed normal subgroup of G . Then the subgroup $[H, G]$, generated (algebraically) by all commutators $[h, g] = h^{-1}g^{-1}hg$ ($h \in H, g \in G$), is closed in G .*

This implies that the (algebraic) derived group $\gamma_2(G) = [G, G]$ is closed, and then by induction that each term $\gamma_n(G) = [\gamma_{n-1}(G), G]$ of the lower central series of G is closed. It is an elementary (though not trivial) fact that $\gamma_n(G)$ is actually the verbal subgroup for the word $\gamma_n(x_1, x_2, \dots, x_n) = [x_1, x_2, \dots, x_n]$.

Theorem 1.5 *Let $q \in \mathbb{N}$ and let G be a finitely generated non-universal profinite group. Then the subgroup G^q is open in G .*

Here G^q denotes the subgroup generated (algebraically) by all q th powers in G , and G is said to be *non-universal* if there exists at least one finite group that is not isomorphic to any open section B/A of G (that is, with $A \triangleleft B \leq G$ and A open in G). We do not know whether this condition is necessary for Theorem 1.5; it seems to be necessary for our proof.

Although the word $w = x^q$ is not in general locally finite, we may still infer that G^q is open once we know that G^q is *closed* in G . The argument is exactly the same as before; far from being a triviality, however, it depends in this case on Zelmanov's theorem [Z] which asserts that there is a finite upper bound $\bar{\mu}(d, q)$ for the order of any *finite* d -generator group of exponent dividing q (the solution of the restricted Burnside problem).

The words γ_n (for $n \geq 2$) are also not locally finite. Could it be that verbal subgroups of finitely generated profinite groups are in general closed? The answer is no: Romankov [R] has constructed a finitely generated (and soluble) pro- p group G in which the second derived group G'' is not closed; and $G'' = w(G)$ where $w = [[x_1, x_2], [x_3, x_4]]$.

Uniform bounds for finite groups

Qualitative statements about profinite groups may often be interpreted as quantitative statements about (families of) finite groups. For example, a profinite group G is finitely generated if and only if there exists a natural number d such that every continuous finite quotient of G can be generated by d elements.

To re-interpret the theorems stated above, consider a group word $w = w(x_1, x_2, \dots, x_k)$. For any group G we write

$$G^{\{w\}} = \{w(g_1, g_2, \dots, g_k)^{\pm 1} \mid g_1, g_2, \dots, g_k \in G\},$$

and call this the set of w -values in G . If the group G is profinite, the mappings $\mathbf{g} \mapsto w(\mathbf{g})$ and $\mathbf{g} \mapsto w(\mathbf{g})^{-1}$ from $G^{(k)}$ to G are continuous, so the set $G^{\{w\}}$ is compact. For any subset S of G let us write

$$S^{*n} = \{s_1 s_2 \dots s_n \mid s_1, \dots, s_n \in S\}.$$

Then for each natural number n , the set $(G^{\{w\}})^{*n}$ of all products of n w -values in G is compact, hence closed in G .

Now $w(G)$ is the ascending union of compact sets

$$w(G) = \bigcup_{n=1}^{\infty} (G^{\{w\}})^{*n}.$$

If $w(G)$ is closed in G , a straightforward application of the Baire category theorem (see [Hr]) shows that for some finite n one has

$$w(G) = (G^{\{w\}})^{*n}. \quad (1)$$

The *converse* (which is more important here) is obvious. Thus $w(G)$ is closed if and only if (1) holds for some natural number n . Now this is a property that can be detected in the finite quotients of G . That is,

- $w(G) = (G^{\{w\}})^{*n}$ if and only if $w(G/N) = ((G/N)^{\{w\}})^{*n}$ for every open normal subgroup N of G .

The “only if” is obvious; to see the other implication, write \mathcal{N} for the set of all open normal subgroups of G and observe that if $w(G/N) = ((G/N)^{\{w\}})^{*n}$ for each $N \in \mathcal{N}$ then

$$w(G) \subseteq \bigcap_{N \in \mathcal{N}} w(G)N = \bigcap_{N \in \mathcal{N}} (G^{\{w\}})^{*n} N = (G^{\{w\}})^{*n}$$

because $(G^{\{w\}})^{*n}$ is closed.

It follows that Theorem 1.3 is equivalent to

Theorem 1.6 *Let d be a natural number and let w be a d -locally finite word. Then there exists $f = f(w, d)$ such that in every finite d -generator group G , every element of the verbal subgroup $w(G)$ is a product of f w -values.*

A similar argument shows that Theorem 1.4 is a consequence of

Theorem 1.7 *Let G be a finite d -generator group and H a normal subgroup of G . Then every element of $[H, G]$ is equal to a product of $g(d)$ commutators $[h, y]$ with $h \in H$ and $y \in G$, where $g(d) = 12d^3 + O(d^2)$ depends only on d .*

In particular, this shows that in any finite d -generator group G , each element of the derived group $\gamma_2(G) = [G, G]$ is equal to a product of $g(d)$ commutators. Now let $n > 2$. It is easy to establish identities of the following type: (a)

$[y_1, \dots, y_n]^{-1} = [y_2, y_1, y'_3, \dots, y'_n]$ where y'_j is a certain conjugate of y_j for $j \geq 3$, and (b) for $k \geq 2$, $[c_1 \dots c_k, x] = [c'_1, x'_1] \dots [c'_k, x'_k]$ where c'_j is conjugate to c_j and x'_j is conjugate to x for each j . Using these and arguing by induction on n we infer that each element of $\gamma_n(G) = [\gamma_{n-1}(G), G]$ is a product of $g(d)^{n-1}$ terms of the form $[y_1, \dots, y_n]$. Thus Theorems 1.6 and 1.7 together imply Theorem 1.2.

For a finite group H let us denote by $\alpha(H)$ the largest integer k such H involves the alternating group $\text{Alt}(k)$ (i.e. such that $\text{Alt}(k) \cong M/N$ for some $N \triangleleft M \leq H$). Evidently, a profinite group G is non-universal if and only if the numbers $\alpha(\tilde{G})$ are bounded as \tilde{G} ranges over all the finite continuous quotients of G , and we see that Theorem 1.5 is equivalent to

Theorem 1.8 *Let q, d and c be natural numbers. Then there exists $h = h(c, d, q)$ such that in every finite d -generator group G with $\alpha(G) \leq c$, every element of G^q is a product of h q th powers.*

It is worth remarking (though not surprising) that the functions f, g and h necessarily depend on the number of generators d (i.e. they must be unbounded as $d \rightarrow \infty$). This can be seen e.g. from the examples constructed by Holt in [Ho], Lemma 2.2: among these are finite groups K (with $\alpha(K) = 5$) such that $K = [K, K] = K^2$ but with $\log |K| / \log |K^{\{w\}}|$ unbounded, where $w(x) = x^2$ (for the application to g note that every commutator is a product of three squares). The Cartesian product G of infinitely many such groups is then a topologically perfect profinite group (i.e. G has no proper open normal subgroup with abelian quotient), but the subgroup G^2 is not closed; in particular $G > G^2$ so G contains a (non-open) subgroup of index 2.

The proofs depend ultimately on two theorems about finite simple groups. We state these here, but postpone their proofs, which rely on the Classification and use quite different methods, to Part II [NS].

Let α, β be automorphisms of a group S . For $x, y \in S$, we define the “twisted commutator”

$$T_{\alpha, \beta}(x, y) = x^{-1} y^{-1} x^{\alpha} y^{\beta},$$

and write $T_{\alpha, \beta}(S, S)$ for the set $\{T_{\alpha, \beta}(x, y) \mid x, y \in S\}$ (in contrast to our convention that $[S, S]$ denotes the group generated by all $[x, y]$). Recall that a group S is said to be *quasisimple* if $S = [S, S]$ and $S/Z(S)$ is simple (here $Z(S)$ denotes the centre of S).

Theorem 1.9 *There is an absolute constant $D \in \mathbb{N}$ such that if S is a finite quasisimple group and $\alpha_1, \beta_1, \dots, \alpha_D, \beta_D$ are any automorphisms of S then*

$$S = T_{\alpha_1, \beta_1}(S, S) \cdot \dots \cdot T_{\alpha_D, \beta_D}(S, S).$$

Theorem 1.10 *Let q be a natural number. There exist natural numbers $C = C(q)$ and $M = M(q)$ such that if S is a finite quasisimple group with $|S/Z(S)| > C$, β_1, \dots, β_M are any automorphisms of S , and q_1, \dots, q_M are any divisors of q , then there exist inner automorphisms $\alpha_1, \dots, \alpha_M$ of S such that*

$$S = [S, (\alpha_1 \beta_1)^{q_1}] \cdot \dots \cdot [S, (\alpha_M \beta_M)^{q_M}].$$

(Here the notation $[S, \gamma]$ stands for the *set* of all $[x, \gamma]$, $x \in S$, not the group they generate.)

Arrangement of the paper

The rest of the paper is devoted to the proofs of Theorems 1.6, 1.7 and 1.8. All groups henceforth will be assumed finite (apart from the occasional appearance of free groups).

In §2 we state what we call the *Key Theorem*, a slightly more elaborate version of Theorem 1.7, and show that it implies Theorem 1.6. Once this is done, we can forget all about the mysterious word w . Section 3 presents two variants of the Key Theorem, and the deduction of Theorems 1.7 and 1.8.

The proof of the Key Theorem is explained in §4. The argument is by induction on the group order, and the inductive step requires a number of subsidiary results. These are established in §§5, 7, 9, 10 and 11, while Sections 6 and 8 contain necessary preliminaries. (To see just the complete proof of Theorem 1.1, the reader may skip §3, the last subsection of §4 and §11.)

Historical remarks

The special cases of Theorems 1.1, 1.4 and 1.5 relating to prosoluble groups were established in [Sg], and the global strategy of our proofs follows the same model.

The special case of Theorem 1.8 where q is odd was the main result of [N1]. Theorem 1.8 for simple groups G (the result in this case being independent of $\alpha(G)$) was obtained by [MZ] and [SW]; a common generalization of this result and of Theorem 1.6 for simple groups is given in [LS2], and is the starting point of our proof. Theorem 1.9 generalizes a result from [W].

The material of Sections 6 and 7 generalizes (and partly simplifies) methods from [Sg], while that of Sections 8-11 extends techniques introduced in [N1] and [N2].

We are indebted to J. S. Wilson for usefully drawing our attention to the verbal subgroup $w(G)$ where w defines the variety generated by a finite group.

Notation

Here G denotes a group, $x \in G$, $y \in G$ or $y \in \text{Aut}(G)$, $S, T \subseteq G$, $q \in \mathbb{N}$.

$$\begin{aligned} x^y &= y^{-1}xy, & [x, y] &= x^{-1}x^y \\ [S, y] &= \{[s, y] \mid s \in S\} \\ \mathfrak{c}(S, T) &= \{[s, t] \mid s \in S, t \in T\} \\ S^{\{q\}} &= \{s^q \mid s \in S\} \\ ST &= \{st \mid s \in S, t \in T\} \\ S^{*q} &= \{s_1 s_2 \dots s_q \mid s_1, \dots, s_q \in S\} \\ &= SS \dots S \text{ (} q \text{ factors),} \end{aligned}$$

and $\langle S \rangle$ denotes the subgroup generated by S . If $H, K \leq G$ (meaning that H and K are subgroups of G),

$$\begin{aligned} [H, K] &= [H, {}_1 K] = \langle \mathfrak{c}(H, K) \rangle, \\ [H, {}_n K] &= [[H, {}_{n-1} K], K] \quad (n > 1), \\ [H, {}_\omega K] &= \bigcap_{n \geq 1} [H, {}_n K], \\ H' &= [H, H]. \end{aligned}$$

The n th Cartesian power of a set S is generally denoted $S^{(n)}$, and n -tuples are conventionally denoted by boldface type: $(s_1, \dots, s_n) = \mathbf{s}$.

$\alpha(G)$ denotes the largest integer k such that G involves the alternating group $\text{Alt}(k)$.

The term ‘simple group’ will mean ‘non-abelian finite simple group’.

2 The Key Theorem

The following theorem is the key to the main results. We make an *ad hoc*

Definition Let H be a normal subgroup of a finite group G . Then H is *acceptable* if

- (i) $H = [H, G]$,
- (ii) if $Z < N$ are normal subgroups of G contained in H then N/Z is neither a (non-abelian) simple group nor the direct product of two isomorphic (non-abelian) simple groups.

Key Theorem Let $G = \langle g_1, \dots, g_d \rangle$ be a finite group and H an acceptable normal subgroup of G . Let q be a natural number. Then

$$H = ([H, g_1] \cdot \dots \cdot [H, g_d])^{*h_1(d, q)} \cdot (H^{\{q\}})^{*z(q)}$$

where $h_1(d, q)$ and $z(q)$ depend only on the indicated arguments.

Assuming this result, let us prove Theorem 1.6. Fix an integer $d \geq 2$ and a group word $w = w(x_1, \dots, x_k)$; we assume that

$$\mu := \mu(d, w) = |F_d/w(F_d)|$$

is finite, where F_d denotes the free group of rank d . Let q denote the order of $C/w(C)$ where C is the infinite cyclic group. Evidently $q \mid \mu$, and it is easy to see that $C^q = w(C) = C^{\{w\}}$; hence

$$h^q \in H^{\{w\}} \tag{2}$$

for any group H and $h \in H$.

Let \mathcal{S} denote the set of simple groups S that satisfy $w(S) = 1$. It follows from the Classification that every simple group can be generated by two elements ; therefore $|\mathcal{S}| \leq \mu(2, w)$ for each $S \in \mathcal{S}$, so the set \mathcal{S} is *finite*. We shall denote the complementary set of simple groups by \mathcal{T} .

An important special case of our theorem was established by Liebeck and Shalev (it is valid for arbitrary words w ; in the present case, it may also be deduced, via (2), from the main result of [MZ] and [SW], together with the fact that there are only finitely many simple groups of exponent dividing q):

Proposition 2.1 ([LS2], Theorem 1.6.) *There exists a constant $c(w)$ such that*

$$S = (S^{\{w\}})^{*c(w)}$$

for every $S \in \mathcal{T}$.

The next result is due to Hamidoune:

Lemma 2.2 [Hm] *Let X be a generating set of a group G such that $1 \in X$ and $|G| \leq r|X|$. Then $G = X^{*2r}$.*

We call a group Q *semisimple* if Q is a direct product of simple groups, and *quasi-semisimple* if $Q = Q'$ and $Q/Z(Q)$ is semisimple. In this case, Q is a central quotient of its universal covering group \tilde{Q} , and \tilde{Q} is a direct product of quasisimple groups.

Corollary 2.3 *Let Q be a quasi-semisimple group having no composition factors in \mathcal{S} . Then*

$$Q = (Q^{\{w\}})^{*n_1}$$

where $n_1 = 2q^2c(w) + q$.

Proof. In view of the preceding remark, we may assume that Q is in fact quasisimple. Write $Z = Z(Q)$ and put $X = Q^{\{w\}}$. It is evident that X generates Q modulo Z ; since $\langle X \rangle \triangleleft Q = Q'$ it follows that X generates Q . According to Proposition 2.1 we have $Q = ZX^{*c}$ where $c = c(w)$.

Now it follows from the Classification (see [G], Table 4.1 or [GLS], §6.1) that Z has rank at most 2. If we assume for the moment that $Z^q = 1$, we may infer that $|Z| \leq q^2$, so $|Q| \leq q^2|X^{*c}|$. In this case, Hamidoune's lemma yields $Q = X^{*2q^2c}$. In general we may conclude that

$$Q = Z^q \cdot X^{*2q^2c},$$

and the result follows since Z is abelian and every q th power is a w -value. ■

Lemma 2.4 *Let G be a group, H a normal subgroup and suppose that $G = G' \langle x_1, \dots, x_m \rangle$. Then*

$$[H, G] = [H, x_1] \dots [H, x_m][H, {}_n G]$$

for every $n \geq 1$.

Proof. Suppose this holds for a certain value of $n \geq 1$. To deduce that it holds with $n+1$ in place of n we may as well assume that $[H_{n+1}G] = 1$. This implies that $[[H_{n-1}G], G'] = 1$. Now $[H_nG]$ is generated by elements of the form $[w, g]$ with $w \in [H_{n-1}G]$ and $g \in G$. As $[H_nG]$ is central in G it follows that every element of $[H_nG]$ takes the form

$$z = [w_1, x_1] \dots [w_m, x_m]$$

with $w_i \in [H_{n-1}G]$ for each i . For any $h_1, \dots, h_m \in H$ we then have

$$\begin{aligned} [h_1, x_1] \dots [h_m, x_m] \cdot z &= [w_1, x_1][h_1, x_1] \dots [w_m, x_m][h_m, x_m] \\ &= [w_1 h_1, x_1] \dots [w_m h_m, x_m], \end{aligned}$$

again because each $[w_i, x_i]$ is central. Thus

$$[H, G] = [H, x_1] \dots [H, x_m][H_nG] = [H, x_1] \dots [H, x_m]$$

as required. ■

Proof of Theorem 1.6. Let G be a d -generator finite group and put $\mathfrak{X} = G^{\{w\}}$. We shall show that

$$w(G) = \mathfrak{X}^{*f}, \quad (3)$$

where $f = f(w, d)$ is a number that will be specified in due course.

We begin by setting up a configuration to which the Key Theorem may be applied. Set

$$\begin{aligned} G_1 &= w(G), \\ H_1 &= \bigcap_{\theta \in \Theta} \ker \theta \end{aligned}$$

where Θ is the set of all homomorphisms from G_1 to $\text{Aut}(S \times S)$ with $S \in \mathcal{S}$. Set

$$H_2 = [H_1, {}_\omega G_1].$$

Then H_1/H_2 is nilpotent and $H_2 = [H_2, G_1]$. Define H_3 to be the smallest normal subgroup of H_1 such that H_1/H_3 is soluble; then $H_3 \leq H_2$ and $H_3 = H'_3$. Set

$$H_4 = \bigcap \mathcal{N}$$

where \mathcal{N} is the set of all normal subgroups K of H_3 such that $H_3/K \in \mathcal{T}$. Finally, put

$$H_5 = [H_4, H_3].$$

Note that H_3/H_4 is a semisimple group; it follows that H_3/H_5 is quasi-semisimple.

Next, we choose a nice generating set for G_1 . Since $F_d/w(F_d)$ is finite, the group $w(F_d)$ is generated by finitely many w -values in F_d :

$$w(F_d) = \langle w(\mathbf{u}_1), \dots, w(\mathbf{u}_{d'}) \rangle.$$

Choose an epimorphism $\pi : F_d \rightarrow G$ and put $g_i = \pi(w(\mathbf{u}_i))$ for $i = 1, \dots, d'$. Then

$$G_1 = w(G) = \langle g_1, \dots, g_{d'} \rangle$$

and for each i we have $g_i = w(\pi(\mathbf{u}_i)) \in \mathfrak{X}$. Note that d' depends only on w and d , and that

$$[h, g_i] = g_i^{-h} g_i \in \mathfrak{X}^{*2} \quad (4)$$

for each i and any $h \in G$.

Now we build up to the proof of (3) in steps.

Step 1. $H_5 \subseteq \mathfrak{X}^{*n_2}$ where $n_2 = 2d'h_1(d', q) + z(q)$. We show first that H_5 is an acceptable subgroup of G_1 . To verify condition (i), observe that $H_5 = [H_5, H_3]$ because $H_3 = H'_3$, so $H_5 = [H_5, G_1]$. For condition (ii), suppose that $Z < N$ are normal subgroups of G_1 contained in H_5 and that $N/Z = S_1 \times \dots \times S_n$ where $n \leq 2$ and the S_j are isomorphic simple groups. If $S_1 \in \mathcal{S}$ then H_1 must act trivially by conjugation on N/Z , which is impossible since $N \leq H_1$ and N/Z is non-abelian. Therefore $S_1 \in \mathcal{T}$. Now H_3 permutes the factors S_j by conjugation, and as $H_3 = H'_3$ and $n \leq 2$ it follows that $S_1 \triangleleft H_3/Z$. Since the outer automorphism group of S_1 is soluble (Schreier's conjecture, [G], the action of H_3 on S_1 induces precisely the group of inner automorphisms of S_1 ; consequently $H_3/C_{H_3}(S_1) \cong S_1$. Hence $C_{H_3}(S_1) \geq H_4 \geq N$, a contradiction since S_1 is non-abelian.

We may now apply the Key Theorem to the pair (G_1, H_5) . This shows that each element of H_5 is equal to one of the form

$$\prod_{j=1}^{h_1(d', q)} \prod_{i=1}^{d'} [a_{ij}, g_i] \cdot \prod_{j=1}^{z(q)} b_j^q,$$

and the claim follows by (4) and (2).

Step 2. $H_3 \subseteq \mathfrak{X}^{*n_1} H_5$ where $n_1 = 2q^2 c(w) + q$. This follows from Corollary 2.3 applied to the quasi-semisimple group H_3/H_5 .

Step 3. $H_2 \subseteq \mathfrak{X}^{*n_2} H_3$. It is clear that H_2/H_3 is an acceptable subgroup of G_1/H_3 . The claim now follows just as in Step 1, on applying the Key Theorem to the pair $(G_1/H_3, H_2/H_3)$.

Step 4. $[H_1, G_1]H_1^q \subseteq \mathfrak{X}^{*(2d'+1)} H_2$. Note that $H_2 = [H_{1,n} G_1]$ for some n ; now Lemma 2.4, with (4), shows that $[H_1, G_1] \subseteq \mathfrak{X}^{*2d'} H_2$, and the claim follows by (2) since $H_1^q \subseteq [H_1, G_1] \cdot H_1^{\{q\}}$.

Step 5. $G_1 \subseteq \mathfrak{X}^{*n_3} [H_1, G_1]H_1^q$ where n_3 depends only on d' and w . Let ν denote the maximal order of $\text{Aut}(S \times S)$ as S ranges over \mathcal{S} (it is easy to see that $\nu \leq 2\mu(2, w)^4$.) For each such S the number of homomorphisms $G_1 \rightarrow \text{Aut}(S \times S)$ is at most $\nu^{d'}$, so $|G_1 : H_1| \leq \nu^{\nu^{d'}} = \rho$, say. It follows that H_1 can be generated by $\rho d'$ elements, and hence that $|H_1 : [H_1, G_1]H_1^q| \leq \rho^{d'}$. Thus $|G_1/[H_1, G_1]H_1^q| \leq n_3$ where $n_3 = \rho^{d'}$; consequently each of its elements can be written as a word of length at most n_3 in the images of the generators g_i .

Conclusion. Putting Steps 1 – 5 together we obtain (3) with

$$f = n_1 + 2n_2 + 2d' + 1 + n_3.$$

3 Variations on a theme

In this section we present two variants of the Key Theorem, and use them to deduce Theorems 1.7 and 1.8. The variants will be proved at the end of §4.

The first variant of the Key Theorem has the same hypotheses, but a new conclusion (its proof will not need Theorem 1.10 or the material of §10).

Key Theorem (B) *Let $G = \langle g_1, \dots, g_d \rangle$ be a finite group and H an acceptable normal subgroup of G . Then*

$$H = ([H, g_1] \cdot \dots \cdot [H, g_d])^{*h_2(d)} \cdot \mathfrak{c}(H, H)^{*D}$$

where $h_2(d) = 6d^2 + O(d)$ depends only on d and D is an absolute constant (given in Theorem 1.9).

Proof of Theorem 1.7. Let $G = \langle g_1, \dots, g_d \rangle$ be a finite group and H a normal subgroup of G . Putting

$$\mathfrak{X} = \mathfrak{c}(H, G)$$

we shall show that

$$[H, G] = \mathfrak{X}^{*g(d)} \tag{5}$$

where $g(d) \leq 2dh_2(d) + O(d)$ is a number that depends only on d . Obviously, if H is acceptable this follows at once from Key Theorem (B), with $g(d) = dh_2(d) + D$. For the general case, we take a step by step approach as in the preceding section.

Put

$$H_1 = [H, {}_\omega G];$$

let H_2 be the smallest normal subgroup of H such that H/H_2 is soluble; let

$$H_3 = \bigcap \mathcal{N}$$

where \mathcal{N} denotes the set of all normal subgroups K of H_2 such that H_2/K is (non-abelian) simple; and put

$$H_4 = [H_3, H_2].$$

As in the preceding section, we see that $H_4 = [H_4, H_2] = [H_4, G]$ and that H_2/H_4 is a quasi-semisimple group. We shall need

Lemma 3.1 *If Q is a quasi-semisimple group then $Q = \mathfrak{c}(Q, Q)^{*D}$.*

Replacing Q by its universal cover, we may suppose that Q is a direct product of quasisimple groups; in that case, the result follows from the special case of Theorem 1.9 where all the automorphisms α_j and β_j are equal to the identity (this special case may be quickly deduced, using Lemma 2.2, from Wilson's theorem [W], Prop. 2.1).

Step 1_B. $H_4 \subseteq \mathfrak{X}^{*(dh_2(d)+D)}$. As remarked above, this holds provided H_4 is an acceptable normal subgroup of G . That this is the case follows, just as in Step 1 of the preceding section, from the fact that H_3 is contained in the kernel of every homomorphism $H_2 \rightarrow \text{Aut}(S \times S)$, S any simple group; the argument is now much simpler since we may ignore the distinction made there between different kinds of simple group.

Step 2_B. $H_2 \subseteq \mathfrak{X}^{*D} H_4$. This follows from Lemma 3.1 applied to the quasi-simple group H_2/H_4 .

Step 3_B. $H_1 \subseteq \mathfrak{X}^{*(dh_2(d)+D)} H_2$. This follows from Key Theorem (B) applied to the pair $(G/H_2, H_1/H_2)$; it is clear that H_1/H_2 is an acceptable normal subgroup of G/H_2 .

Step 4_B. $[H, G] \subseteq \mathfrak{X}^{*d} H_1$. This is immediate from Lemma 2.4.

Conclusion. Putting the steps together we obtain (5) with

$$g(d) = 2dh_2(d) + 3D + d = 12d^3 + O(d^2).$$

■

The second variant of the Key Theorem has a weaker hypothesis: as we shall see, this is necessary because the failure of the word $w(x) = x^q$ to be locally finite means that we have less control over the generators of the verbal subgroup G^q . (The proof of this variant will not need Theorem 1.10 or the material of §§10, 11.)

Key Theorem (C) *Let G be a d -generator finite group and H an acceptable normal subgroup of G . Suppose that $G = H \langle g_1, \dots, g_r \rangle$. Then*

$$H = ([H, g_1] \cdot \dots \cdot [H, g_r])^{*h_3(d, c)}$$

where $h_3(d, c)$ depends only on d and $c = \alpha(G)$.

Proof of Theorem 1.8. Let G be a d -generator group with $\alpha(G) \leq c$, let q be a natural number, and put $\mathfrak{X} = G^{\{q\}}$. We will prove that

$$G^q = \mathfrak{X}^{*h}, \tag{6}$$

where $h = h(c, d, q)$ will be determined below.

To this end, we take $w(x) = x^q$ and then define $G_1 = G^q$ and normal subgroups $H_1 \geq \dots \geq H_5$ exactly as in the proof of Theorem 1.6 in §2. The argument now follows that proof step by step, but we have to carry out the

steps in reverse order: this is necessary in order to obtain substitutes for the ‘global generators’ g_i used in §2.

As in the preceding section, we will repeatedly use the fact that if $h \in G$ and $g \in \mathfrak{X}$ then $[h, g] \in \mathfrak{X}^{*2}$.

Set

$$\mu = \overline{\mu}(d, q),$$

the maximal order of a finite d -generator group of exponent dividing q ; this is finite by the positive solution of the restricted Burnside problem [Z]. Then $|G : G_1| \leq \mu$, and it follows that G_1 can be generated by $d' = d\mu$ elements.

Since G_1 is generated by \mathfrak{X} , the argument of §2, Step 5 now gives

Step 5_C. $G_1 \subseteq \mathfrak{X}^{*n_3}[H_1, G_1]H_1^q$ where n_3 depends only on d' and q .

The next step depends on the following simple observation, where $\sigma(q)$ will denote the number of distinct prime divisors of q .

Lemma 3.2 *If $H = \langle X \rangle$ is an r -generator abelian group then*

$$H = \langle y_1^q, \dots, y_r^q, x_1, \dots, x_{r\sigma(q)} \rangle$$

for some $y_1, \dots, y_r \in H$ and some $x_1, \dots, x_{r\sigma(q)} \in X$.

Proof. Let P be a Sylow p -subgroup of H . Write $\pi : H \rightarrow P$ for the projection. P is an r -generator p -group generated by $\pi(X)$ so $P = \langle \pi(X_p) \rangle$ for some subset X_p of X of size r (because $P/\text{Frat}(P)$ is an r -dimensional \mathbb{F}_p -vector space). Thus if p_1, \dots, p_σ are the primes dividing q and P_1, \dots, P_σ the corresponding Sylow subgroups, then the subgroup $R = \langle X_{p_1} \cup \dots \cup X_{p_\sigma} \rangle$ projects onto each P_i . It follows that $|H : R|$ is coprime to q and hence that $H = QR$ where Q is a direct factor of H of order coprime to q . Thus Q is an r -generator group and each element of Q is a q th power, so $Q = \langle y_1^q, \dots, y_r^q \rangle$ for some y_1, \dots, y_r . The lemma follows. ■

Applying this lemma to G_1/G_1' , we deduce that

$$G_1 = G_1' \langle h_1, \dots, h_{d''} \rangle$$

where each $h_i \in \mathfrak{X}$ and $d'' = d'(1 + \sigma(q))$. Now Lemma 2.4 gives

$$[H_1, G_1] = \prod_{i=1}^{d''} [H_1, h_i] \cdot H_2 \subseteq \mathfrak{X}^{*2d''} H_2.$$

As $H_1^q \subseteq [H_1, G_1]H_1^{\{q\}}$ we have established

Step 4_C. $[H_1, G_1]H_1^q \subseteq \mathfrak{X}^{*(2d''+1)} H_2$.

Putting the last two steps together gives $G_1 = \mathfrak{X}^{*n_4} H_2$ where n_4 depends only on d and q . As G_1 is generated by d' elements, it follows that there exist $g_1, \dots, g_r \in \mathfrak{X}$, where $r = n_4 d'$, such that

$$G_1 = H_2 \langle g_1, \dots, g_r \rangle.$$

Since H_2/H_3 is an acceptable normal subgroup of G_1/H_3 , Key Theorem (C) may be applied to give

Step 3_C. $H_2 \subseteq \mathfrak{X}^{*n_5} H_3$ where $n_5 = 2rh_3(d', c)$.

Step 2_C. $H_3 \subseteq \mathfrak{X}^{*n_1} H_5$ where n_1 depends only on q . This is identical to Step 2 in §2.

Step 1_C. $H_5 \subseteq \mathfrak{X}^{*n_2}$ where n_2 depends only on q, d and c . We proved in Step 1 of §2 that H_5 is an acceptable normal subgroup of G_1 . So the claim will follow by Key Theorem (C) if we can show that $G_1 = H_5 \langle g'_1, \dots, g'_s \rangle$ where each $g'_i \in \mathfrak{X}$ and s depends only on q, d and c . But this follows from the preceding four steps: for G_1 is generated by d' elements, each of which lies in $\mathfrak{X}^{*n_6} H_5$ where $n_6 = n_1 + n_5 + n_4$, so we may take $s = d'n_6$.

Conclusion. Altogether we obtain (6) with $h = n_6 + n_2$.

4 Proof of the Key Theorem

The general idea

Before getting down to specifics, let us outline the general plan of attack. The Key Theorem asserts that, under suitable hypotheses on the finite group G and its normal subgroup H , every element of H is equal to a product of a specific form. Thus what has to be established is the solvability of equations like

$$h = \Phi(u_1, \dots, u_m) \quad (7)$$

where

$$\Phi(u_1, \dots, u_m) = U(g_1, \dots, g_r, u_1, \dots, u_m);$$

here the ‘constant’ h is an arbitrary element of H , U is a specific group word, g_1, \dots, g_r are some fixed parameters from G , and the ‘unknowns’ u_1, \dots, u_m are to be found in H . The idea of the proof is modelled on that of Hensel’s Lemma: one shows that an approximate solution of (7) can be successively refined to an exact solution.

What makes Hensel’s Lemma work is a hypothesis that ensures the surjectivity of a certain linear map: the relevant derivative must be non-singular modulo p . This translates in a straightforward way to our context.

Definition Let $\mathbf{v} \in H^{(m)}$. The mapping $\Phi'_{\mathbf{v}} : H^{(m)} \rightarrow H$ is defined by

$$\Phi(\mathbf{x} \cdot \mathbf{v}) = \Phi'_{\mathbf{v}}(\mathbf{x}) \cdot \Phi(\mathbf{v}) \quad (\mathbf{x} \in H^{(m)})$$

where $\mathbf{x} \cdot \mathbf{v}$ denotes the m -tuple $(x_1 v_1, \dots, x_m v_m)$.

Suppose now that K is a normal subgroup of G contained in H , and that we have found a solution of (7) modulo K ; that is, we have $\mathbf{v} \in H^{(m)}$ such that

$$h = \kappa \cdot \Phi(\mathbf{v})$$

for some $\kappa \in K$. Then $\mathbf{u} = \mathbf{x} \cdot \mathbf{v}$ is a solution of (7) if and only if

$$\Phi'_{\mathbf{v}}(\mathbf{x}) = \kappa. \quad (8)$$

Thus our ‘approximate solution’ \mathbf{v} can be lifted to an exact solution provided the image of the map $\Phi'_{\mathbf{v}}$ contains K . Let us call \mathbf{v} ‘liftable’ in this case. To ensure that the process can be iterated, however, we require that the ‘new’ solution $\mathbf{x} \cdot \mathbf{v}$ is again liftable in the appropriate sense. This will be achieved by a ‘probabilistic’ argument: we establish independently (a) that a relatively large proportion of the elements \mathbf{x} in a suitable domain are solutions of (8), and (b) that a relatively large proportion of the \mathbf{x} in the same domain have the property that $\mathbf{x} \cdot \mathbf{v}$ is liftable. It will follow that at least some of these elements \mathbf{x} will have both properties.

Here is a final remark. All our main results about finite groups concern functions that are uniformly bounded in terms of d , the number of generators. Why is this the dominant parameter? There are two reasons. The first is evident in the statement of the Key Theorem: each of the d generators appears explicitly in the statement. The second, hidden in the proof, is to do with the way the generators have to act on chief factors of the group; it comes down to the following obvious but crucial observation:

Lemma 4.1 *Let $G = \langle g_1, \dots, g_d \rangle$ be a group.*

- (i) *If G acts without fixed points on a set of size n then at least one of the g_i moves at least n/d points.*
- (ii) *If G acts linearly on a vector space V of dimension n , and fixes only 0, then at least one of the g_i satisfies $\dim C_V(g_i) \leq (1 - \frac{1}{d})n$.*

(Here $C_V(g)$ denotes the fixed-point set of g .)

Solvability of equations

Let G be a finite group. A normal subgroup N of G will be called *quasi-minimal* if $N = [N, G] > 1$ and N is minimal with this property. It is easy to see that in this case, there is a uniquely determined normal subgroup $Z = Z_N$ of G maximal subject to $Z < N$; indeed, if Z_1 and Z_2 were two distinct such subgroups then $N = Z_1 Z_2$ would imply $[N, G] = [N, Z_1][N, Z_2] = 1$.

We write ‘QMN’ for ‘quasi-minimal normal subgroup’, and recall the definition of ‘acceptable’ from §3. The Frattini subgroup of G is denoted $\text{Frat}(G)$.

Lemma 4.2 *Let N be a QMN of G and put $Z = Z_N$. Suppose that $N \leq H$ where H is an acceptable normal subgroup of G . Then*

- (i) *N/Z is a minimal normal subgroup of G/Z , $[Z, {}_k G] = 1$ for some k , and $[Z, N] \leq [Z, H] = 1$.*
- (ii) *$Z \leq \text{Frat}(G)$.*
- (iii) *If N is not soluble then N is quasi-semisimple with centre Z and $N/Z = S_1 \times \dots \times S_n$, where $n \geq 3$ and S_1, \dots, S_n are isomorphic non-abelian simple groups.*

(iv) If N is soluble then N/Z is an elementary abelian p -group for some prime p ; also $N^p = 1$ if p is odd, $N^2 = [N, N]$ and $[N, N]^2 = 1$ if $p = 2$.

Proof. (i) The first two statements are immediate from the definition. To show that $[Z, H] = 1$, write $Z_i = [Z, {}_iG]$ for $i \geq 0$ (with $Z_0 = Z$). Then $Z_k = 1$. Suppose we have $[Z_i, H] = 1$ for some i with $k \geq i > 0$. Since $H = [H, G]$ the Three-Subgroup Lemma gives

$$[Z_{i-1}, H] = [[Z_{i-1}, H], G][Z_i, H] = [[Z_{i-1}, H], G]$$

whence $[Z_{i-1}, H] = 1$ since $[Z_{i-1}, H] < N$. It follows by reverse induction that $[Z, H] = [Z_0, H] = 1$.

(ii) Suppose that M is a maximal subgroup of G and M contains Z_i but not Z_{i-1} , where $i > 0$. Then $G = Z_{i-1}M$ and $H = Z_{i-1}(H \cap M)$ so $[H, G] \leq M$, a contradiction since $Z_{i-1} \leq H = [H, G]$.

(iii) This follows from the well-known structure of minimal normal subgroups; here $n \geq 3$ because N is contained in the acceptable subgroup H .

(iv) The first claim is standard. Since $[N, N] \leq Z(N)$, the map $x \mapsto x^p$ is a homomorphism of G -operator groups from N into Z if p is odd, and induces such a homomorphism from N into $Z/[N, N]$ if $p = 2$. In each case the image of this homomorphism must be 1 since $N = [N, G]$. The final statement is easy. ■

The solvability of equations like (8) is assured by the following results, which will be proved in later sections (the fourth one, Proposition 11.1, is needed only for variant (B) of the Key Theorem). In each case, N denotes a QMN of G and $Z = Z_N$.

For $\mathbf{x} = (x_1, \dots, x_t)$, $\mathbf{y} = (y_1, \dots, y_t) \in G^{(t)}$ we will write

$$[\mathbf{x}, \mathbf{y}] = \prod_{j=1}^t [x_j, y_j].$$

Proposition 7.1 *Suppose that N is soluble and that $[Z, G] = 1$. Put $K = N$ if N is abelian, $K = N'$ otherwise. For $i = 1, 2, 3$ define $\phi_i : N^{(m)} \rightarrow N$ by*

$$\phi_i(\mathbf{a}) = [\mathbf{a}, \mathbf{y}_i]$$

where $\mathbf{y}_i = (y_{i1}, \dots, y_{im})$ and the y_{ij} are elements of G such that $\langle y_{i1}, \dots, y_{im} \rangle K = G$ for each i . Let $\kappa \in K$. Then there exist $\kappa_1, \kappa_2, \kappa_3 \in N$ such that $\kappa_1 \kappa_2 \kappa_3 = \kappa$ and, for each $i = 1, 2, 3$,

$$|\phi_i^{-1}(\kappa_i)| \geq |N|^m |N/Z|^{-d-1}.$$

The corresponding results for a non-soluble QMN involve certain constants:

$D \geq 1$ is the absolute constant specified in Theorem 1.9, and we set $\overline{D} = 4 + 2D$;

$C(q)$ and $M(q)$ are the constants specified in Theorem 1.10, and we set $z(q) = M(q)\overline{D}(q + \overline{D})$.

Definition Let $\varepsilon > 0$ and $k \in \mathbb{N}$. Let $\mathbf{y} = (y_1, y_2, \dots, y_m) \in G^{(m)}$.

- (i) The m -tuple \mathbf{y} has the (k, ε) *fixed-point property* if in any transitive permutation action of G on a set of size $n \geq 2$, at least k of the elements y_i move at least εn points.
- (ii) The m -tuple \mathbf{y} has the (k, ε) *fixed-space property* if for every irreducible $\mathbb{F}_p G$ -module V of dimension $n \geq 2$, where p is any prime, at least k of the y_i satisfy $\dim_{\mathbb{F}_p} C_V(y_i) \leq (1 - \varepsilon)n$.

Proposition 9.2 *Suppose that N is quasi-semisimple, and that N/Z is not simple. Define $\phi : N^{(m)} \rightarrow N$ by*

$$\phi(\mathbf{a}) = [\mathbf{a}, \mathbf{y}]$$

where y_1, \dots, y_m are elements of G such that $\langle y_1, \dots, y_m \rangle N = G$. Suppose that \mathbf{y} has the (k, ε) fixed-point property where $k\varepsilon \geq \overline{D}$. Then for each $\kappa \in N$,

$$|\phi^{-1}(\kappa)| \geq |N|^m |N/Z|^{-4D}.$$

Proposition 10.1 *Let $q \in \mathbb{N}$. Suppose that N is quasi-semisimple, and that its non-abelian composition factors S satisfy $|S| > C(q)$. Let $u_1, \dots, u_m \in G$ where $m \geq z(q)$. Then the mapping $\psi : N^{(m)} \rightarrow N$ defined by*

$$\prod_{j=1}^m (x_j u_j)^q = \psi(\mathbf{x}) \prod_{j=1}^m u_j^q$$

is surjective.

Proposition 11.1 *Suppose that N is quasi-semisimple, and let $\alpha_1, \beta_1, \dots, \alpha_D, \beta_D$ be $2D$ arbitrary automorphisms of N . Then the mapping $\theta : N^{(2D)} \rightarrow N$ defined by*

$$\theta(\mathbf{a}, \mathbf{b}) = \prod_{j=1}^D T_{\alpha_j, \beta_j}(a_j, b_j)$$

is surjective.

Lifting generators

The other half of our probabilistic argument rests on the following proposition, which will be established in §5. For a simple group S we define $\mu(S)$ to be the supremum of the numbers μ such that

$$|S : M| \geq |S|^\mu$$

for every maximal subgroup M of S , and for any group N define

$$\mu'(N) = \min \left\{ \frac{1}{2}, \mu(S) \mid S \text{ a non-abelian composition factor of } N \right\}.$$

For later use, we also define

$$\mu(q) = \min \left\{ \frac{1}{2}, \mu(S) \mid S \text{ simple, } |S| \leq C(q) \right\}.$$

Proposition 5.1 *Let G be a d -generator group and N an acceptable QMN of G . Suppose that $G = \langle y_1, \dots, y_m \rangle N$. Put $Z = Z_N$ and let*

$$\mathcal{N}(\mathbf{y}) = \left\{ \mathbf{a} \in N^{(m)} \mid \langle y_1^{a_1}, \dots, y_m^{a_m} \rangle \neq G \right\}.$$

Let $\varepsilon \in (0, \frac{1}{2}]$.

(i) *Suppose that N is soluble and that \mathbf{y} has the (k, ε) fixed-space property. Then*

$$|\mathcal{N}(\mathbf{y})| \leq |N|^m |N/Z|^{d-k\varepsilon}.$$

(ii) *There exists an absolute constant C_0 such that if N is quasi-semisimple and \mathbf{y} has the (k, ε) fixed-point property, where $k\varepsilon \geq \max\{2d + 4, 2C_0 + 2\}$, then*

$$\begin{aligned} |\mathcal{N}(\mathbf{y})| &< |N|^m \quad \text{and} \\ |\mathcal{N}(\mathbf{y})| &\leq |N|^m |N/Z|^{1-s} \end{aligned}$$

where

$$s = \min\{\mu'(N)(k\varepsilon/2 - d - 1), \mu'(N)(k\varepsilon/2 - C_0)\}.$$

The proof

Now we can prove the Key Theorem, assuming the results stated above. We will need to know the following ‘derivative’, obtained by direct calculation:

Lemma 4.3 *Let $\mathbf{g} \in G^{(m)}$. Define $\Xi : G^{(m)} \rightarrow G$ by $\Xi(\mathbf{v}) = [\mathbf{v}, \mathbf{g}]$. Then*

$$\Xi'_{\mathbf{v}}(\mathbf{x}) = \prod_{j=1}^m [x_j, g_j]^{\tau_j(\mathbf{g}, \mathbf{v})}$$

where

$$\tau_j(\mathbf{g}, \mathbf{v}) = v_j [g_{j-1}, v_{j-1}] \dots [g_1, v_1].$$

Now define

$$k(d, q) = 1 + \left\lceil d \cdot \max \left\{ \frac{8D+2}{\mu(q)} + 2d + 2, \frac{8D+2}{\mu(q)} + 2C_0 \right\} \right\rceil$$

(where $\lceil x \rceil$ denotes the least integer $\geq x$), and let $z(q)$ be as defined above. The first claim in the next proposition gives the Key Theorem, on putting

$$h_1(d, q) = 3k(d, q).$$

Proposition 4.4 *Let $G = \langle g_1, \dots, g_d \rangle$ and let H be an acceptable normal subgroup of G . Let $m = d \cdot k(d, q)$ and define $\mathbf{g} = (g_1, \dots, g_m)$ by setting*

$$g_{td+i} = g_i \quad (0 \leq t < k(d, q)).$$

Then for each $h \in H$ there exist $\mathbf{v}(1), \mathbf{v}(2), \mathbf{v}(3) \in H^{(m)}$ and $\mathbf{u} \in H^{(z(q))}$ such that

$$h = \prod_{i=1}^3 [\mathbf{v}(i), \mathbf{g}] \cdot \prod_{l=1}^{z(q)} u_l^q \quad (9)$$

and

$$\langle g_1^{\tau_1(\mathbf{g}, \mathbf{v}(i))}, \dots, g_m^{\tau_m(\mathbf{g}, \mathbf{v}(i))} \rangle = G \quad \text{for } i = 1, 2, 3. \quad (10)$$

The second claim, (10), is required for the inductive proof. In terms of the heuristic discussion above, it ensures that our solution $(\mathbf{v}(1), \mathbf{v}(2), \mathbf{v}(3), \mathbf{u})$ is again ‘liftable’: in the guise of (17) or (18), it is used directly in ‘Case 1’, below, and in other cases enables us to quote some of the above-stated propositions, whose hypotheses stipulate that a certain set of elements should generate an appropriate quotient of G .

Let us recall that H is *acceptable* in G if (i) $H = [H, G]$ and (ii) no normal section of G inside H takes the form S or $S \times S$ for a non-abelian simple group S . It is clear that H/K is then acceptable in G/K whenever $H \geq K$ and $K \triangleleft G$; we shall use this without special mention.

Proof. We will write $k = k(d, q)$ and $z = z(q)$. The result is trivial if $H = 1$; we suppose that $H > 1$ and argue by induction on $|H|$. Since $H = [H, G]$ it follows that H contains a QMN N of G . It also follows that $d \geq 2$. Put $Z = Z_N$ and define a normal subgroup $K > 1$ of G as follows:

$$K = \begin{cases} [Z, G] & \text{if } [Z, G] > 1 \\ N & \text{if } [Z, G] = 1 \text{ and } [N, N] = 1 \\ [N, N] & \text{if } [Z, G] = 1 \text{ and } [N, N] > 1 \end{cases} \quad (11)$$

Write the equation (9) as

$$h = \Phi(\mathbf{v}, \mathbf{u}) = \Xi(\mathbf{v}(1)) \cdot \Xi(\mathbf{v}(2)) \cdot \Xi(\mathbf{v}(3)) \cdot \Psi(\mathbf{u}).$$

Inductively, we may assume that there exist $\kappa \in K$, $\mathbf{v}(i) \in H^{(m)}$ and $\mathbf{u} \in H^{(z)}$ such that

$$h = \kappa \Phi(\mathbf{v}, \mathbf{u})$$

and, for $i = 1, 2, 3$,

$$\left\langle g_1^{\tau_1(\mathbf{v}(i))}, \dots, g_m^{\tau_m(\mathbf{v}(i))} \right\rangle K = G, \quad (12)$$

where for brevity we write $\tau_j(\mathbf{x}) = \tau_j(\mathbf{g}, \mathbf{x})$.

The aim is to show that there exist $\mathbf{a}(i) \in N^{(m)}$ and $\mathbf{b} \in N^{(z)}$ such that (9) and (10) hold with $\mathbf{a}(i) \cdot \mathbf{v}(i)$ replacing $\mathbf{v}(i)$ and $\mathbf{b} \cdot \mathbf{u}$ replacing \mathbf{u} . The first requirement is equivalent to

$$\begin{aligned} \kappa &= \Phi'_{(\mathbf{v}, \mathbf{u})}(\mathbf{a}, \mathbf{b}) \\ &= \Xi'_{\mathbf{v}(1)}(\mathbf{a}(1))^{\xi_1} \cdot \Xi'_{\mathbf{v}(2)}(\mathbf{a}(2))^{\xi_2} \cdot \Xi'_{\mathbf{v}(3)}(\mathbf{a}(3))^{\xi_3} \cdot \Psi'_{\mathbf{u}}(\mathbf{b})^{\xi_4} \end{aligned} \quad (13)$$

where $\xi_1 = 1$ and

$$\xi_i = (\Xi(\mathbf{v}(1)) \dots \Xi(\mathbf{v}(i-1)))^{-1} \quad (i = 2, 3, 4).$$

It is convenient to reformulate the second requirement. Write

$$\bar{a}(i)_j = a(i)_j^{v(i)_j g_j \dots g_m}, \quad \bar{g}_{ij} = g_j^{v(i)_j g_j \dots g_m}.$$

Lemma 4.5 *Let $\mathbf{a}(i) \in N^{(m)}$ for $i = 1, 2, 3$. The following are equivalent, for each i :*

$$G = \left\langle g_1^{\tau_1(\mathbf{a}(i) \cdot \mathbf{v}(i))}, \dots, g_m^{\tau_m(\mathbf{a}(i) \cdot \mathbf{v}(i))} \right\rangle, \quad (14)$$

$$G = Z \left\langle \bar{g}_{i1}^{\bar{a}(i)_1}, \dots, \bar{g}_{im}^{\bar{a}(i)_m} \right\rangle. \quad (15)$$

Proof. We claim that for any m -tuple \mathbf{v} ,

$$\left\langle g_1^{\tau_1(\mathbf{v})}, \dots, g_m^{\tau_m(\mathbf{v})} \right\rangle^{g_1 \dots g_m} = \langle g_1^{v_1 g_1 \dots g_m}, \dots, g_m^{v_m g_m} \rangle. \quad (16)$$

To see this, put $z_1 = 1$ and for $k > 1$ set

$$z_k = g_{k-1}^{v_{k-1} g_{k-2}^{-1} \dots g_1^{-1}} \cdot z_{k-1}.$$

Arguing by induction on k we find that $z_{k+1} = z_k g_k^{\tau_k(\mathbf{v})}$ for each k ; this implies that

$$\left\langle g_1^{\tau_1(\mathbf{v})}, \dots, g_m^{\tau_m(\mathbf{v})} \right\rangle = \langle z_2, \dots, z_{m+1} \rangle = \left\langle g_1^{v_1}, \dots, g_m^{v_m g_{m-1}^{-1} \dots g_1^{-1}} \right\rangle$$

which is equivalent to (16).

The lemma follows on taking $\mathbf{v} = \mathbf{a}(i) \cdot \mathbf{v}(i)$, and noting that $\bar{g}_{ij}^{\bar{a}(i)_j} = g_j^{a(i)_j v(i)_j g_j \dots g_m}$ and $Z \leq \text{Frat}(G)$. ■

Taking each $a(i)_j = 1$ and replacing G by G/K , we deduce that (12) implies

$$G = \langle \bar{g}_{i1}, \dots, \bar{g}_{im} \rangle K \quad (i = 1, 2, 3). \quad (17)$$

Now write

$$\tilde{g}_{ij} = g_j^{\tau_j(\mathbf{v}(i))\xi_i}, \quad \tilde{a}(i)_j = a(i)_j^{\tau_j(\mathbf{v}(i))\xi_i}.$$

Then (12) is also (evidently) equivalent to

$$G = \langle \tilde{g}_{i1}, \dots, \tilde{g}_{im} \rangle K \quad (i = 1, 2, 3); \quad (18)$$

and Lemma 4.3 shows that

$$\Xi'_{\mathbf{v}(i)}(\mathbf{a}(i))^{\xi_i} = [\tilde{\mathbf{a}}(i), \tilde{\mathbf{g}}_i] \quad (i = 1, 2, 3). \quad (19)$$

Thus it suffices to find $\mathbf{a}(i)$ and \mathbf{b} (with entries in N) such that

$$\kappa = [\tilde{\mathbf{a}}(1), \tilde{\mathbf{g}}_1][\tilde{\mathbf{a}}(2), \tilde{\mathbf{g}}_2][\tilde{\mathbf{a}}(3), \tilde{\mathbf{g}}_3]\Psi'_{\mathbf{u}}(\mathbf{b})^{\xi_4} \quad (20)$$

and such that (15) holds. To this end we separate several cases.

Case 1: where $[Z, G] = K > 1$. We think of Z as a G -module, with K acting trivially, and write it additively. From (18) we have

$$K = Z(G - 1) = \sum_{j=1}^m Z(\tilde{g}_{1j} - 1) = \left\{ [\mathbf{z}, \tilde{\mathbf{g}}_1] \mid \mathbf{z} \in Z^{(m)} \right\}.$$

Thus there exists $\mathbf{a}(1) \in Z^{(m)}$ with $[\mathbf{a}(1), \tilde{\mathbf{g}}_1] = \kappa$, and we may satisfy (20) by setting $a(2)_j = a(3)_j = b_j = 1$ for all j ; note that $\tilde{\mathbf{a}}(1) = \mathbf{a}(1)$ here since $[Z, H] = 1$. As each $\tilde{a}(i)_j$ is in Z and $K \leq Z$, in this case (15) follows at once from (17). ■

Assume henceforth that $[Z, G] = 1$. For $\kappa \in N$ and $1 \leq i \leq 3$ put

$$\mathfrak{X}_i(\kappa) = \left\{ \mathbf{a}(i) \in N^{(m)} \mid [\tilde{\mathbf{a}}(i), \tilde{\mathbf{g}}_i] = \kappa \right\},$$

and let

$$\mathfrak{Y}_i = \left\{ \mathbf{a}(i) \in N^{(m)} \mid \left\langle \tilde{g}_{i1}^{\tilde{a}(i)_1}, \dots, \tilde{g}_{im}^{\tilde{a}(i)_m} \right\rangle Z = G \right\}.$$

We shall repeatedly use the following

Key Observation: For each $i = 1, 2, 3$, the m -tuple $\tilde{\mathbf{g}}_i$ has the $(k, \frac{1}{d})$ fixed-space property and the $(k, \frac{1}{d})$ fixed-point property.

Indeed, since $G = \langle g_1, \dots, g_d \rangle$, Lemma 4.1 shows that the d -tuple (g_1, \dots, g_d) has the $(1, \frac{1}{d})$ fixed-space property and the $(1, \frac{1}{d})$ fixed-point property. The claim follows because each of the generators g_l ($1 \leq l \leq d$) is conjugate to at least k of the elements \tilde{g}_{ij} ($1 \leq j \leq m$).

Case 2: where N is soluble, and $K = N$ if N is abelian, $K = N'$ if not. Define $\phi_i : N^{(m)} \rightarrow N$ by

$$\phi_i(\mathbf{x}) = [\mathbf{x}, \tilde{\mathbf{g}}_i].$$

In view of (18), we may take $y_{ij} = \tilde{g}_{ij}$ in Proposition 7.1 and infer that there exist $\kappa_1, \kappa_2, \kappa_3 \in N$ with $\kappa_1 \kappa_2 \kappa_3 = \kappa$ such that

$$|\mathfrak{X}_i(\kappa_i)| = |\phi_i^{-1}(\kappa_i)| \geq |N|^m |N/Z|^{-d-1}$$

for $i = 1, 2, 3$ (the first equality holds because $\mathbf{a}(i) \mapsto \tilde{\mathbf{a}}(i)$ is a bijection on $N^{(m)}$).

Let $i \in \{1, 2, 3\}$. With the *Key Observation* and (17), Proposition 5.1(i) shows that the number of elements $\mathbf{x} \in N^{(m)}$ for which

$$\langle \bar{g}_{i1}^{x_1}, \dots, \bar{g}_{im}^{x_m} \rangle \neq G$$

is at most $|N|^m |N/Z|^{d-k/d}$. Since $\mathbf{a}(i) \mapsto \bar{\mathbf{a}}(i)$ is a bijection on $N^{(m)}$ this gives

$$|N^{(m)} \setminus \mathfrak{Y}_i| \leq |N|^m |N/Z|^{d-k/d}.$$

As $k > d(2d+1)$, it follows that $|\mathfrak{X}_i(\kappa_i)| > |N^{(m)} \setminus \mathfrak{Y}_i|$.

Thus we may choose $\mathbf{a}(i) \in \mathfrak{X}_i(\kappa_i) \cap \mathfrak{Y}_i$, for $i = 1, 2, 3$. Then (15) holds and (20) is satisfied with $b_l = 1$ for all l . ■

Case 3: where $N = K$ is quasi-semisimple and $|S| \leq C(q)$; here S denotes the (unique) non-abelian composition factor of N .

Put $\kappa_1 = \kappa, \kappa_2 = \kappa_3 = 1$. Using Proposition 9.2 in place of Proposition 7.1, we see just as in Case 2 that for $i = 1, 2, 3$,

$$|\mathfrak{X}_i(\kappa_i)| \geq |N|^m |N/Z|^{-4D};$$

note that $k/d > \bar{D}$ because $D \geq 1 > \mu(q)$.

Now $k/d > \max\{2d+4, 2C_0+2\}$, so Proposition 5.1(ii), with the *Key Observation* and (17), shows that

$$|N^{(m)} \setminus \mathfrak{Y}_i| \leq |N|^m |N/Z|^{1-s}$$

for each i , where

$$\begin{aligned} s &= \min\{\mu(q)(k/2d - d - 1), \mu(q)(k/2d - C_0)\} \\ &> 4D + 1. \end{aligned}$$

We conclude as in the preceding case that (20) and (15) can be simultaneously satisfied by a suitable choice of $\mathbf{a}(1), \mathbf{a}(2), \mathbf{a}(3) \in N^{(m)}$, taking each $b_l = 1$. ■

Case 4: where $N = K$ is quasi-semisimple and $|S| > C(q)$. Applying Proposition 5.1(ii) again we infer that each of the sets \mathfrak{Y}_i is non-empty. Choose $\mathbf{a}(i) \in \mathfrak{Y}_i$ for $i = 1, 2, 3$. Then (15) holds.

Now Proposition 10.1 shows that the mapping $\Psi'_u : N^{(z)} \rightarrow N$ is surjective. Hence there exists $\mathbf{b} \in N^{(z)}$ such that

$$\Psi'_u(\mathbf{b}) = \left(([\tilde{\mathbf{a}}(1), \tilde{\mathbf{g}}_1][\tilde{\mathbf{a}}(2), \tilde{\mathbf{g}}_2][\tilde{\mathbf{a}}(3), \tilde{\mathbf{g}}_3])^{-1} \kappa \right)^{\xi_4^{-1}}.$$

Then (20) is satisfied, and the proof is complete. ■

Remark. It may be worth observing that in Case 3, the only role played by the upper bound on $|S|$ is to provide the lower bound $\mu(q)$ for $\mu(S)$. In fact such a lower bound will obtain if we allow S to range, additionally, over groups of Lie type with bounded Lie ranks (but over finite fields of arbitrary size); this follows from Lemma 4.8, below, for example. We may therefore, if we prefer, restrict Case 4 to where S is either alternating of large degree or of Lie type with large Lie rank. This means that for the Key Theorem, only the special case of Proposition 10.1 relating to such simple groups S is actually needed. This in turn depends only on the corresponding special case of Theorem 1.10; thus (for present purposes) one can do without the fair-sized chunk of Part II devoted to the proof of Theorem 1.10 for groups of Lie type with small Lie rank over large fields. (However, for groups of this type we shall still need the rather easier special case of Theorem 1.10 where $q = 1$, in order to deduce Theorem 1.9.)

Variants (B) and (C)

Define

$$k(d) = 1 + \lceil d \cdot \max\{2d + 4, 2C_0 + 2\} \rceil.$$

Now modify the statement of Proposition 4.4 as follows: replace $k(d, q)$ by $k(d)$, replace $z(q)$ by $2D$, and replace the formula (9) by

$$h = \prod_{i=1}^3 [\mathbf{v}(i), \mathbf{g}] \cdot \prod_{l=1}^D [u_l, u_{l+D}]. \quad (21)$$

This gives Key Theorem (B) if we set $h_2(d) = 3k(d)$.

For the proof of the modified proposition, we set $\Psi(\mathbf{x}, \mathbf{y}) = \prod_{j=1}^D [x_j, y_j]$ and use

Lemma 4.6

$$\Psi'_{(\mathbf{u}, \mathbf{w})}(\mathbf{x}, \mathbf{y}) = \prod_{l=1}^D T_{\alpha_l, \beta_l}(x_l^{\sigma_l}, y_l^{\rho_l})$$

where $\alpha_l, \beta_l, \sigma_l$ and ρ_l are given by certain group words in $u_1, w_1, \dots, u_D, w_D$.

This is verified by direct calculation. We now argue exactly as before, with the following changes: omit Case 3 altogether; and in Case 4, remove the restriction on $|S|$ and use Proposition 11.1 in place of Proposition 10.1. With Lemma 4.6, this shows that the relevant mapping $\Psi'_{\mathbf{u}} : N^{(2D)} \rightarrow N$ is surjective. ■

The modifications required for Key Theorem (C) are a little more drastic, so let us state the appropriate variant of Proposition 4.4. Define

$$k'(d, c) = 1 + \left\lceil d \cdot \max \left\{ \frac{8D + 2}{\varepsilon(c)} + 2d + 2, \frac{8D + 2}{\varepsilon(c)} + 2C_0 \right\} \right\rceil$$

where $\varepsilon(c)$ is the constant appearing in Lemma 4.8 below.

Proposition 4.7 *Let G be a d -generator group with $\alpha(G) = c$ and let H be an acceptable normal subgroup of G . Suppose that $G = H \langle g_1, \dots, g_r \rangle$. Put $m = r \cdot k'(d, c)$ and define $\mathbf{g} = (g_1, \dots, g_m)$ by setting*

$$g_{tr+i} = g_i \quad (0 \leq t < k'(d, c)).$$

Then for each $h \in H$ there exist $\mathbf{v}(1), \mathbf{v}(2), \mathbf{v}(3) \in H^{(m)}$ such that

$$h = \prod_{i=1}^3 [\mathbf{v}(i), \mathbf{g}]$$

and

$$\langle g_1^{\tau_1(\mathbf{g}, \mathbf{v}(i))}, \dots, g_m^{\tau_m(\mathbf{g}, \mathbf{v}(i))} \rangle = G \quad \text{for } i = 1, 2, 3.$$

Key Theorem (C) then follows on setting $h_3(d, c) = 3k'(d, c)$. For the proof, we may no longer appeal to Lemma 4.1; instead we rely on

Lemma 4.8 *There exists $\varepsilon = \varepsilon(c) \in (0, \frac{1}{2}]$, depending only on $c = \alpha(G)$, such that the following hold.*

(i) *If G acts as a primitive permutation group on a set Ω of size ≥ 2 , with kernel G_Ω , then $|\Omega| \geq |G : G_\Omega|^\varepsilon$.*

(ii) *For each transitive G -set Ω of size ≥ 2 , there is a proper normal subgroup $G_0(\Omega)$ of G such that for each $x \in G \setminus G_0(\Omega)$,*

$$|\text{fix}_\Omega(x)| \leq (1 - \varepsilon) |\Omega|$$

(where $\text{fix}_\Omega(x)$ denotes the set of fixed points of x in Ω).

(iii) *For each simple $\mathbb{F}_p G$ -module V there is a proper normal subgroup $G_0(V)$ of G such that for each $x \in G \setminus G_0(V)$,*

$$\dim C_V(x) \leq (1 - \varepsilon) \dim V.$$

Proof. Gluck, Seress and Shalev prove in [GSS], Theorem 1.2 that every primitive G -set Ω contains a *base* B of size at most $\gamma = \gamma(c)$, a number depending only on c (to say that B is a base means that the pointwise stabilizer of B is equal to G_Ω). We may suppose that $\gamma \geq 2$. This gives (i) with $\varepsilon = \gamma^{-1}$, since the action of each element of G is determined by where it moves each element of B . (In fact (i) is a celebrated result of Babai, Cameron and Pálffy [BCP].)

It also implies (ii) for the case of a primitive action. To see this, let $x \in G \setminus G_\Omega$, let $\omega \in \Omega$ and put $X = \{y \in G \mid \omega x^y \neq \omega\}$. Then

$$Xg_1 \cup \dots \cup Xg_\gamma = G$$

where $B = \{\omega g_1, \dots, \omega g_\gamma\}$ so $|X| \geq \gamma^{-1} |G|$. Therefore $\Omega \setminus \text{fix}_\Omega(x) = \{\omega y^{-1} \mid y \in X\}$ has cardinality at least $\gamma^{-1} |G| / |G_\omega| = \gamma^{-1} |\Omega|$. In this case (ii) follows with $\varepsilon = \gamma^{-1}$ and $G_0(\Omega) = G_\Omega$.

The general case of (ii) follows on taking $G_0(\Omega)$ to be the kernel of the induced action on a minimal system of imprimitivity.

Statement (iii) for a primitive $\mathbb{F}_p G$ -module V is Theorem 5.3 of [GSS], with $G_0(V) = C_G(V)$. When V is imprimitive, take $G_0(V)$ to be the kernel of the permutation action of G on a minimal system of imprimitivity in V , and apply (ii). (A better bound for $\gamma(c)$ is given in [LS1], Theorem 1.4.) ■

The proof now proceeds as in the preceding subsection, simply omitting the function Ψ . The *Key Observation* is replaced by

Key Observation (C). Let $k = k'(d, c)$. For each $i = 1, 2, 3$, the image in $(G/K)^{(m)}$ of the m -tuple $\bar{\mathbf{g}}_i$ has the (k, ε) fixed-space property and the (k, ε) fixed-point property.

To see this, recall (17), which asserts that the $\bar{g}_{ij}K$ ($j = 1, \dots, m$) generate G/K . Lemma 4.8(ii) then implies that for any transitive G/K -set of size $n \geq 2$, at least one of the elements $\bar{g}_{ij}K$ must move at least εn points. Since each \bar{g}_{ij} is conjugate to at least k of the \bar{g}_{il} this shows that $(\bar{g}_{i1}K, \dots, \bar{g}_{im}K)$ has the (k, ε) fixed-point property. The (k, ε) fixed-space property follows likewise from Lemma 4.8(iii).

Now the Key Observation is applied in conjunction with Propositions 9.2 and 5.1. Both of these only really need the relevant ‘ (k, ε) -hypothesis’ to be satisfied by the image of the m -tuple \mathbf{y} in $(G/N)^{(m)}$ (see §5 and §9). As $K \leq N$ this means that we may use Key Observation (C) just as we used the Key Observation in the preceding subsection.

Cases 1, 2. Exactly as before, replacing $1/d$ by ε where necessary.

Case 3: where $N = K$ is quasi-semisimple. Let S denote the (unique) non-abelian composition factor of N . Then Lemma 4.8(i) shows that $|S : M| \geq |S|^\varepsilon$ for each maximal subgroup M of S , so we have $\mu(N) \geq \varepsilon$. The argument then proceeds as before, with ε in place of $\mu(q)$. ■

5 The first inequality: lifting generators

In this section, we fix a finite d -generator group G and an acceptable quasi-minimal normal subgroup N of G . Thus N contains a normal subgroup Z of G with $Z \leq \text{Frat}(G)$ such that N/Z is a minimal normal subgroup of G/Z , and if N/Z is non-abelian then N/Z is not the product of fewer than 3 simple groups.

In the latter case, the composition factors of N/Z are all isomorphic to a simple group S , and we have defined $\mu(S)$ to be the supremum of the numbers τ such that

$$|S : M| \geq |S|^\tau$$

for every maximal subgroup M of S . We will write $\mu = \min\{\mu(S), \frac{1}{2}\}$.

Fix positive integers k and m and let $\varepsilon > 0$. Recall the

Definition Let $\mathbf{y} = (y_1, y_2, \dots, y_m) \in G^{(m)}$.

- (i) The m -tuple \mathbf{y} has the (k, ε) *fixed-point property* if in any transitive permutation action of G on a set of size $n \geq 2$, at least k of the elements y_i move at least εn points.

- (ii) The m -tuple \mathbf{y} has the (k, ε) *fixed-space property* if for every irreducible $\mathbb{F}_p G$ -module V of dimension $n \geq 2$, where p is any prime, at least k of the y_i satisfy $\dim_{\mathbb{F}_p} C_V(y_i) \leq (1 - \varepsilon)n$.

We shall prove

Proposition 5.1 *Let $y_1, \dots, y_m \in G$ and assume that $G = \langle y_1, \dots, y_m \rangle N$. Put*

$$\mathcal{N}(\mathbf{y}) = \left\{ \mathbf{a} \in N^{(m)} \mid \langle y_1^{a_1}, \dots, y_m^{a_m} \rangle \neq G \right\}.$$

Let $\varepsilon \in (0, \frac{1}{2}]$.

- (i) *Suppose that N is soluble and that \mathbf{y} has the (k, ε) fixed-space property. Then*

$$|\mathcal{N}(\mathbf{y})| \leq |N|^m |N/Z|^{d-k\varepsilon}.$$

- (ii) *There exists an absolute constant C_0 such that if N is quasi-semisimple and \mathbf{y} has the (k, ε) fixed-point property, where $k\varepsilon \geq \max\{2d + 4, 2C_0 + 2\}$, then*

$$\begin{aligned} |\mathcal{N}(\mathbf{y})| &< |N|^m \quad \text{and} \\ |\mathcal{N}(\mathbf{y})| &\leq |N|^m |N/Z|^{1-s} \end{aligned}$$

where

$$s = \min\{\mu(k\varepsilon/2 - d - 1), \mu(k\varepsilon/2 - C_0)\}.$$

In fact, in (i) the fixed-space property of \mathbf{y} will only be applied to the action of G on the elementary abelian group N/Z , and in (ii) the fixed-point property of \mathbf{y} will only be applied to the permutation action of G on the simple factors of N/Z ; so in both cases it would be enough to assume that the relevant property is possessed by the image of \mathbf{y} in $(G/N)^{(m)}$. (This is used in the proof of Key Theorem (C).)

For $\mathbf{a} \in N^{(m)}$ write

$$Y(\mathbf{a}) = \langle y_1^{a_1}, \dots, y_m^{a_m} \rangle,$$

so $\mathcal{N}(\mathbf{y}) = \{\mathbf{a} \in N^{(m)} \mid Y(\mathbf{a}) \neq G\}$. Since $Z \leq \text{Frat}(G)$ we have

$$Y(\mathbf{a}) \neq G \iff Y(\mathbf{a})Z \neq G,$$

so $\mathcal{N}(\mathbf{y})$ is the union of a certain number r , say, of cosets of $Z^{(m)}$. If we show that $r \leq |N/Z|^{m-t}$ it will follow that $|\mathcal{N}(\mathbf{y})| \leq |Z|^m |N/Z|^{m-t} = |N|^m |N/Z|^{-t}$. Thus we may replace G by G/Z and so assume henceforth that $Z = 1$.

We now proceed with the proof. If N is soluble then it is a simple $\mathbb{F}_p G$ -module for some prime p , so in case (i) at least k of the y_i satisfy $\dim C_N(y_i) \leq (1 - \varepsilon)n$ where $n = \dim N$. If N is not soluble then $N = S_1 \times \dots \times S_n$ where $n \geq 3$ and G permutes the set $\Omega = \{S_1, \dots, S_n\}$ transitively by conjugation; so in case (ii) at least k of the y_i move at least εn of the factors S_j ; for each such

i we have $|C_N(y_i)| \leq |N|^{1-\varepsilon/2}$ (cf. Lemma 5.5 below). Thus in either case, we may relabel the y_i so that

$$|C_N(y_i)| \leq |N|^{1-\bar{\varepsilon}} \quad \text{for } 1 \leq i \leq k \quad (22)$$

where

$$\bar{\varepsilon} = \begin{cases} \varepsilon & (N \text{ soluble}) \\ \varepsilon/2 & (N \text{ insoluble}) \end{cases}.$$

Now if $\mathbf{a} \in \mathcal{N}(\mathbf{y})$ then $Y(\mathbf{a}) \leq M$ for some maximal subgroup M of G with $NM = G$. Write

$$v(M; y) = |\{a \in N \mid y^a \in M\}|$$

and

$$v(M) = \prod_{i=1}^m v(M; y_i).$$

Then $v(M)$ is just the number of \mathbf{a} such that $Y(\mathbf{a}) \leq M$, so

$$|\mathcal{N}(\mathbf{y})| \leq \sum_{M \in \mathcal{M}} v(M) \quad (23)$$

where \mathcal{M} denotes the set of maximal subgroups of G which supplement N .

Lemma 5.2 *Let $M \in \mathcal{M}$, $y \in G$ and put $D = M \cap N$. Then*

$$v(M; y) = |C_N(y)| \cdot |[y^b, N] \cap D|$$

for every $b \in N$ such that $y^b \in M$, and $v(M; y) = 0$ if there is no such b .

Proof. If no conjugate of y lies in M then $v(M; y) = 0$. Otherwise, $y^b \in M$ for some $b \in N$; given any such b , for $a \in N$ we have

$$y^a \in M \iff [y^b, b^{-1}a] \in M \cap [y^b, N] = [y^b, N] \cap D.$$

The lemma follows since the fibres of the mapping $a \mapsto [y^b, b^{-1}a]$ are cosets of $C_N(y)$. ■

Let \mathcal{M}_0 denote the set of all $M \in \mathcal{M}$ such that $M \cap N = 1$.

Lemma 5.3 (i)

$$|\mathcal{M}_0| \leq |N|^d.$$

(ii) *If $M \in \mathcal{M}_0$ then*

$$v(M) \leq |N|^{m-k\bar{\varepsilon}}.$$

Proof. (i) Follows from the well-known fact that the complements to N in G , if there are any, correspond bijectively to derivations from G/N to N , and the fact that G can be generated by d elements.

(ii) Since $N \cap M = 1$, Lemma 5.2 and (22) give

$$v(M) = \prod_{j=1}^m v(M; y_j) \leq |N|^{k(1-\bar{\varepsilon})} \cdot |N|^{m-k} = |N|^{m-k\bar{\varepsilon}}.$$

■

Part (i) of the proposition now follows: for when N is abelian we have $\mathcal{M} = \mathcal{M}_0$ and so

$$|\mathcal{N}(\mathbf{y})| \leq |N|^d \cdot |N|^{m-k\varepsilon}$$

as required.

We assume henceforth that N is *non-abelian*; thus

$$N = S_1 \times \cdots \times S_n$$

where $n \geq 3$ and the S_i are isomorphic simple groups. The conjugation action of G permutes the factors S_i transitively, and we write

$$S_i^g = S_{i\sigma(g)}$$

where $\sigma(g) \in \text{Sym}(n)$.

For a natural number e put

$$\mathcal{M}(e) = \{M \in \mathcal{M} \mid |G : M| = e\}.$$

Thus $\mathcal{M}(|N|) = \mathcal{M}_0$, and $\mathcal{M}(e)$ is non-empty only when $e \geq 2$ and e is a divisor of $|N|$.

Lemma 5.4 *There is an absolute constant C such that*

$$|\mathcal{M}(e)| \leq e^C$$

for every proper divisor e of $|N|$.

Proof. Let $M \in \mathcal{M}(e)$ and put $D = M \cap N$. Since $|N : D| = |G : M| = e$ we have $1 < D < N$, so D is not normal in G . As $D \triangleleft M$ it follows that $M \geq C_G(N)$. It is now clear that $C_G(N)$ is the core of M , that is, the biggest normal subgroup of G contained in M .

Thus $M \mapsto M/C_G(N)$ maps $\mathcal{M}(e)$ bijectively onto the set of core-free maximal subgroups in $G/C_G(N)$ that supplement but do not complement $NC_G(N)/C_G(N)$ and have index e . It is proved by Mann and Shalev in [MS] that the cardinality of this set is bounded by e^C where C is an absolute constant: see the first part of the proof of [MS], Corollary 2. ■

Lemma 5.5 *Let $V = A_1 \times \cdots \times A_t$ where $t \geq 2$ and the A_i are isomorphic finite groups. Let g be an automorphism of V that permutes the subgroups A_i and moves at least εt of them.*

(i) *Let $U = B_1 \times \cdots \times B_t$ where $B_i < A_i$ and $|B_i| = |B_1|$ for each i . Suppose that $U^g = U$. Then*

$$|C_V(g)| \cdot |[g, V] \cap U| \leq |V| \cdot |V : U|^{-\varepsilon/2}.$$

(ii) *Let $\Delta \cong A_1$ be a diagonal subgroup of V . Suppose that $\Delta^g = \Delta$ and that $t \geq 3$. Then*

$$|C_V(g)| \cdot |[g, V] \cap \Delta| \leq |V| \cdot |V : \Delta|^{-\varepsilon/2}.$$

Proof. Write $A = A_1$, $B = B_1$. Consider a typical cycle for the permutation action of g , say $\mathcal{C} = (A_1, \dots, A_l)$, and put $V_{\mathcal{C}} = A_1 \times \cdots \times A_l$. Note that for $1 \leq i \leq l$ we have

$$B_i = U \cap A_1^{g^{i-1}} = (U \cap A_1)^{g^{i-1}} = B_1^{g^{i-1}}.$$

A typical element of $V_{\mathcal{C}}$ takes the form

$$v = a_1 \cdot a_2^g \cdot \dots \cdot a_l^{g^{l-1}}$$

where $a_i \in A_1$ for each i . Then

$$[g, v] = a_l^{-g^l} a_1 \cdot (a_1^{-1} a_2)^g \cdot \dots \cdot (a_{l-1}^{-1} a_l)^{g^{l-1}} \quad (24)$$

so $[g, v] \in U$ if and only if

$$\begin{aligned} a_i^{-1} a_{i+1} &\in B_1 \quad (1 \leq i \leq l-1) \\ \theta(a_l)^{-1} a_1 &\in B_1 \end{aligned}$$

where θ denotes the automorphism induced on A_1 by g^l . Hence putting

$$\begin{aligned} X &= \{v \in V_{\mathcal{C}} \mid [g, v] \in U\}, \\ Y &= \{y \in A_1 \mid \theta(y)^{-1} y \in B_1\} \end{aligned}$$

we obtain a bijection

$$\begin{aligned} B_1^{(l-1)} \times Y &\rightarrow X \\ (b_1, \dots, b_{l-1}, y) &\mapsto (\theta(y)b_1) \cdot (\theta(y)b_2)^g \cdot \dots \cdot (\theta(y)b_{l-1})^{g^{l-2}} \cdot y^{g^{l-1}}. \end{aligned}$$

Since the fibres of the map $v \mapsto [g, v]$ are cosets of $C_{V_{\mathcal{C}}}(g)$ it follows that

$$\begin{aligned} |[g, V_{\mathcal{C}}] \cap U| \cdot |C_{V_{\mathcal{C}}}(g)| &= |X| \\ &= |B_1|^{l-1} |Y| \leq |B|^{l-1} |A|. \end{aligned}$$

Now V as a $\langle g \rangle$ -operator group is the direct product of the $V_{\mathcal{C}}$ over all the cycles $\mathcal{C} = \mathcal{C}_1, \dots, \mathcal{C}_p$ say. It follows that

$$\begin{aligned} |[g, V] \cap U| \cdot |C_V(g)| &= \prod_{i=1}^p |[g, V_{\mathcal{C}_i}] \cap U| \cdot |C_{V_{\mathcal{C}_i}}(g)| \\ &\leq \prod_{i=1}^p \left(|B|^{l_i-1} |A| \right) = |A|^t |A : B|^{p-t} \end{aligned}$$

where l_i is the length of \mathcal{C}_i . Since at most $(1 - \varepsilon)t$ of the l_i are equal to 1 we have $p - t \leq -\varepsilon t/2$. Hence

$$|[g, V] \cap U| \cdot |C_V(g)| \leq |A|^t |A : B|^{-\varepsilon t/2} = |V| |V : U|^{-\varepsilon/2}$$

and (i) is proved.

Taking $U = 1$ in (i) we deduce that $|C_V(g)| \leq |A|^p$. Since $|\Delta| = |A|$ it follows that

$$|[g, V] \cap \Delta| \cdot |C_V(g)| \leq |A|^{1+p}.$$

Suppose first that $l_i \geq 2$ for each i . Then $p \leq t/2$, and as $t \geq 3$ we have

$$1 + p \leq t - (t - 1)/4.$$

It follows that $|A|^{1+p} \leq |V| |V : \Delta|^{-1/4}$, and (ii) follows since $\varepsilon \leq \frac{1}{2}$.

Now suppose that one of the l_i is equal to 1, say $l_1 = 1$. Then g fixes A_1 . Each element of $[g, V] \cap \Delta$ is determined by its first component, which belongs to $[g, A_1]$. Applying part (i) to $V^* = A_2 \times \dots \times A_t$, with each $B_i = 1$, we deduce as above that $|C_{V^*}(g)| \leq |A|^{p-1}$ and hence that

$$|C_V(g)| = |C_{A_1}(g)| |C_{V^*}(g)| \leq |C_{A_1}(g)| |A|^{p-1}.$$

It follows that

$$|[g, V] \cap \Delta| \cdot |C_V(g)| \leq |[g, A_1]| |C_{A_1}(g)| |A|^{p-1} = |A|^p.$$

As $p - t \leq -\varepsilon t/2 < -\varepsilon(t - 1)/2$ we have $|A|^p < |V| |V : \Delta|^{-\varepsilon/2}$, again giving (ii). ■

Now fix e with $2 \leq e < |N|$ and consider $M \in \mathcal{M}(e)$. Put $D = M \cap N$ and let R_i denote the projection of D into S_i . It is easy to see that if $g \in M$ then $R_i^g = R_{i\sigma(g)}$ for each i , so the group $\tilde{R} = R_1 \times \dots \times R_n$ is normalized by M .

Say M is of *type 1* if $\tilde{R}M \neq G$. In this case $D = \tilde{R}$. Put $t = n$, $A_i = S_i$ and $B_i = R_i$. Note that

$$e = |N : D| = |S_1 : R_1|^n \geq |S_1|^{\mu n} = |N|^\mu.$$

Now suppose that $\tilde{R}M = G$. Then $\tilde{R} = N$ and D is a subdirect product in $N = S_1 \times \dots \times S_n$. In this case, we can re-label the S_i so that

$$D = B_1 \times \dots \times B_{t'}$$

where $t' \mid n$ and for each i , B_i is a diagonal subgroup of $S_{r(i-1)+1} \times \cdots \times S_{r(i-1)+r}$ with $r = n/t' \geq 2$ ([Cm], Exercise 4.3). If $t' \geq 2$ say M is of *type 2*, and put $t = t'$, $A_i = S_{r(i-1)+1} \times \cdots \times S_{r(i-1)+r}$.

If $t' = 1$ say M is of *type 3*, put $t = n$, take $A_i = S_i$ for each i and put $\Delta = D$.

Again, we have

$$e = |N : D| = |S_1|^{n-t'} = |N|^{1-r^{-1}} \geq |N|^{1/2} \geq |N|^\mu.$$

In each case, the action of M permutes the A_i transitively. Since $G = NM$ it follows that G also permutes the A_i transitively. Writing \mathcal{J} to denote the set of subscripts $l \leq m$ such that y_l moves at least εt of the A_i , we have $|\mathcal{J}| \geq k$ by the (k, ε) fixed-point property of \mathbf{y} .

Now let $l \in \mathcal{J}$. According to Lemma 5.2, if no N -conjugate of y_l lies in M then $v(M; y_l) = 0$; while if $y_l^b \in M$ where $b \in N$ then

$$v(M; y_l) = |C_N(y_l)| \cdot |[y_l^b, N] \cap M|.$$

Put $g = y_l^b$. Then g also moves at least εt of the A_i . Putting $V = N$, and $U = D$ when M is of types 1 or 2, we may apply Lemma 5.5 to deduce that

$$\begin{aligned} |C_N(g)| \cdot |[g, N] \cap D| &\leq |N| \cdot |N : D|^{-\varepsilon/2} \\ &= e^{-\varepsilon/2} |N|. \end{aligned}$$

As $|C_N(g)| = |C_N(y_l)|$ this shows that

$$v(M; y_l) \leq e^{-\varepsilon/2} |N|.$$

Hence

$$v(M) = \prod_{i=1}^m v(M; y_i) \leq \left(e^{-\varepsilon/2} |N| \right)^{|\mathcal{J}|} \cdot |N|^{m-|\mathcal{J}|} \leq e^{-k\varepsilon/2} |N|^m.$$

This holds for each $M \in \mathcal{M}(e)$. With Lemmas 5.3 and 5.4 it gives

$$\begin{aligned} |\mathcal{N}(\mathbf{y})| &\leq \sum_{M \in \mathcal{M}} v(M) \\ &\leq |N|^{m+d-k\varepsilon/2} + \sum_e e^{C-k\varepsilon/2} |N|^m, \end{aligned}$$

where e ranges over integers lying between $|N|^\mu$ and $|N|/2$.

We can now deduce part (ii) of Proposition 5.1. Take $C_0 = C + 1$, and assume that

$$k\varepsilon \geq \max\{2d + 4, 2C_0 + 2\}.$$

Put $t = \min\{k\varepsilon/2 - d, k\varepsilon/2 - C\}$, write $\nu = |N|$ and let ζ denote the Riemann zeta function. Then $t \geq 2$, so

$$|N|^{-m} |\mathcal{N}(\mathbf{y})| \leq \sum_{e \geq 2} e^{-t} \leq \zeta(2) - 1 < 1.$$

This establishes the first claim. For the second, observe that $\zeta(t) < 2 < \nu$, and so

$$|N|^{-m} |\mathcal{N}(\mathbf{y})| \leq \sum_{e \geq \nu^\mu} e^{-t} \leq \zeta(t) \nu^{-\mu(t-1)} \leq \nu^{1-s}$$

where

$$s = \mu(t-1) = \min \{ \mu(k\varepsilon/2 - d - 1), \mu(k\varepsilon/2 - C_0) \}.$$

6 Exterior squares and quadratic maps

In the following section we are going to prove Proposition 7.1. This concerns the solution of certain equations in a soluble quasi-minimal normal subgroup N of a finite group G . When N is *abelian* (‘Case 1’) the result is very easy. When N is *non-abelian*, the problem comes down to studying the fibres over $N' = [N, N]$ of certain mappings ϕ_i from $(N/Z)^{(m)}$ into N (induced by commutation with certain elements of G); here $Z = C_N(G)$, N/Z is a simple $\mathbb{F}_p G$ -module for some prime p , and N' is an \mathbb{F}_p -module contained in Z . If $|N'| = 2$ (‘Case 2’) it turns out that the restriction of each ϕ_i to $\phi_i^{-1}(N') = V_i$ is a quadratic form over \mathbb{F}_2 , and the required result follows from some elementary number theory over \mathbb{F}_2 . The hardest case (‘Case 3’) is when $|N'| > 2$. The mappings $\phi_i|_{V_i}$ are still quadratic polynomial mappings over \mathbb{F}_p , but we may no longer suppose that their co-domain N' is one-dimensional over \mathbb{F}_p , and higher-dimensional algebraic geometry does not deliver the result.

To get round this difficulty, we would like to think of N' as a one-dimensional space over a larger field. Such a structure does not arise naturally, in general; however, N' is an epimorphic image of the exterior square of N/Z , and it was shown in [Sg] that the latter does naturally have the structure of a one-dimensional space over a certain field. This is the key to the main result of this section, Proposition 6.2, which in turn will serve to complete case 3 of the proof of Proposition 7.1.

When p is odd, everything needed for the proof essentially appears in [Sg]; but the proof given in that paper for the ‘even’ case depends crucially on a global solubility assumption, not available to us here, and a new approach is required. In fact we shall deal in a uniform way with the ‘odd’ and ‘even’ cases, by strengthening the method used for the ‘odd’ part in [Sg] (and the very tricky material of §§8 and 9 of [Sg] may now be consigned to a historical footnote).

We need to recall some material from [Sg], §4. Let \overline{G} be a group (assumed finite in [Sg], but this is not necessary) and $R = \mathbb{Z}\overline{G}$ the group ring. Let M be a finite simple right R -module, so M is an $\mathbb{F}_p \overline{G}$ -module for some prime p . We may consider M as an R -bimodule via

$$gu = ug^{-1} \quad (u \in M, g \in \overline{G}),$$

and so define $M \otimes_R M$ and the exterior square

$$\wedge_R^2 M \cong \left(\wedge_{\mathbb{F}_p}^2 M \right) / \left(\wedge_{\mathbb{F}_p}^2 M \right) (\overline{G} - 1)$$

(where \overline{G} acts diagonally on $\wedge_{\mathbb{F}_p}^2 M$). We fix a generator \overline{c} for M and put

$$I = \text{ann}_R(\overline{c})$$

$$S_0 = \{r \in R \mid rI \subseteq I\}.$$

The ring S_0/I may be identified with the finite field $\text{End}_R(M)$ via $s + I \mapsto \widehat{s}$ where

$$\widehat{s}(\overline{c}r) = \overline{c}sr \quad (s \in S_0, r \in R).$$

Suppose now that M admits a non-zero \overline{G} -invariant alternating \mathbb{F}_p -bilinear form. According to Proposition 4.4 of [Sg], there exists a subfield $k = S/I \subseteq S_0/I$ such that for each $s \in S$, $a, b \in M$,

$$\widehat{s}a \otimes b = a \otimes \widehat{s}b$$

holds in $M \otimes_R M$, and such that the induced action of k on $\wedge_R^2 M$ makes $\wedge_R^2 M$ into a 1-dimensional vector space over k . Moreover, $\dim_k(M) \geq 2$, and the mapping

$$(a, b) \mapsto a \wedge b$$

from $M \times M$ to $\wedge_R^2 M$ is k -bilinear.

We shall consider M and $\wedge_R^2 M$ as left S -modules via $S \rightarrow k$.

Now assume that we are given a group G and a normal subgroup B such that $G/B = \overline{G}$. Let A/B be a minimal normal subgroup of \overline{G} such that $A/B = M$ as a \overline{G} -module via conjugation. Assume also that

$$[B, A] = [A', G] = 1,$$

$$|A' : A' \cap [B, G]| > 2$$

and that the mapping $aB \wedge bB \mapsto [a, b]$ ($a, b \in A$) induces an isomorphism

$$\wedge_R^2 M \rightarrow A'. \tag{25}$$

These hypotheses imply that M does admit a non-zero \overline{G} -invariant alternating \mathbb{F}_p -bilinear form: there exists an epimorphism $\theta : A' \rightarrow \mathbb{F}_p$ and then $(aB, bB) \mapsto \theta([a, b])$ is such a form. We may therefore identify A' with the one-dimensional k -space $\wedge_R^2 M$ via (25), and will use additive and multiplicative notation interchangeably for the group operation there. Note that

$$|k| = |A'| > 2,$$

$$|M| = |k|^{\dim_k M} \geq |A'|^2.$$

Fix $c \in A$ such that $cB = \overline{c}$, the chosen generator of M . Suppose that \overline{G} can be generated by d elements.

Proposition 6.1 *Let $x_1, \dots, x_m \in G$ satisfy $B \langle x_1, \dots, x_m \rangle = G$. Then there exist (a) a k -subspace U of $M^{(m)}$, (b) a k -quadratic map $\Phi : U \rightarrow A'$, and (c) for each $\mathbf{z} = (z_1, \dots, z_m) \in A^{(m)}$, a k -linear map $\alpha^{\mathbf{z}} : U \rightarrow A'$ such that*

- (i) $\dim_k U \geq (m - d)\dim_k M$
- (ii) *for each $u \in U$ there exist $a_1, \dots, a_m \in A$ with $(a_1 B, \dots, a_m B) = u$ such that*

$$\Phi(u) + \alpha^{\mathbf{z}}(u) = \left(\prod_{j=1}^m [z_j a_j, x_j] \right) \cdot \left(\prod_{j=1}^m [z_j, x_j] \right)^{-1}.$$

Moreover $\alpha^{(1, \dots, 1)} = \alpha$ is surjective.

Before proving this let us deduce its primary application:

Proposition 6.2 *Let $* : G \rightarrow G^*$ be an epimorphism with $\ker(*) \leq B$ and $[B^*, G^*] = 1$. For $i = 1, 2, 3$ let $x_{i1}, \dots, x_{im} \in G^*$ satisfy $B^* \langle x_{i1}, \dots, x_{im} \rangle = G^*$, and define*

$$\phi_i : A^{*(m)} \rightarrow A^*$$

by

$$\phi_i(a_1, \dots, a_m) = \prod_{j=1}^m [a_j, x_{ij}].$$

Then for each $\kappa \in (A^*)'$ there exist $\kappa_1, \kappa_2, \kappa_3 \in A^*$ such that

$$\kappa_1 \kappa_2 \kappa_3 = \kappa \tag{26}$$

and

$$\phi_i^{-1}(\kappa_i) \text{ contains at least } |M|^{m-d-1} \text{ cosets of } B^{*(m)} \quad (i = 1, 2, 3), \tag{27}$$

provided in case $p = 2$ that $A^* = [A^*, G^*]$ and $(A^*)^2 = (A^*)'$.

Proof. Let \tilde{x}_{ij} denote a preimage in G of $x_{ij} \in G^*$, and let $\Phi_i, \alpha_i : U_i \rightarrow A'$ be the mappings corresponding to $(\tilde{x}_{i1}, \dots, \tilde{x}_{im})$ provided in Proposition 6.1. Let $\tilde{\kappa} \in A'$ be a preimage of κ . Write $\Phi_i + \alpha_i = f_i$. Note that f_i is not the zero map, because α_i is surjective and $|k| > 2$, which implies that a non-zero map cannot be both linear and quadratic over k ; and that for each $u \in U_i$ there exist $a_1, \dots, a_m \in A$ such that $u = (a_1 B, \dots, a_m B)$ and

$$f_i(u)^* = \phi_i(a_1^*, \dots, a_m^*). \tag{28}$$

Since $[B^*, G^*] = 1$, this then holds for every m -tuple (a_1, \dots, a_m) with $(a_1 B, \dots, a_m B) = u$. Similarly, if $\mathbf{z} = (z_1, \dots, z_m) \in A^{(m)}$, $\alpha_i^{\mathbf{z}}$ are as given in Proposition 6.1 and $f_i^{\mathbf{z}} = \Phi_i + \alpha_i^{\mathbf{z}}$, then

$$f_i^{\mathbf{z}}(u)^* = \phi_i(z_1^* a_1^*, \dots, z_m^* a_m^*) \phi_i(z_1^*, \dots, z_m^*)^{-1} \tag{29}$$

whenever $(a_1B, \dots, a_mB) = u$.

Case 1: where $p \neq 2$. First we pick κ_3 . The fibres of the map $f_3 : U_3 \rightarrow A'$ have average size at least $|U_3|/|A'| > |M|^{m-d-1}$, so there exists $\tilde{\kappa}_3 \in A'$ with $|f_3^{-1}(\tilde{\kappa}_3)| > |M|^{m-d-1}$. Now put $\kappa_3 = \tilde{\kappa}_3^*$. Then (28) implies that $\phi_3^{-1}(\kappa_3)$ contains at least $|M|^{m-d-1}$ cosets of $B^{*(m)}$.

Next, let $\tilde{\kappa}_4$ be a preimage of $\kappa_3\kappa_3^{-1}$. According to Lemma 5.1 of [Sg] there exist elements $\tilde{\kappa}_1, \tilde{\kappa}_2 \in A'$ such that $\tilde{\kappa}_1 + \tilde{\kappa}_2 = \tilde{\kappa}_4$ and

$$\begin{aligned} |f_i^{-1}(\tilde{\kappa}_i)| &\geq |k|^{\dim_k U_i - 2} \\ &\geq |M|^{m-d} |k|^{-2} \geq |M|^{m-d-1} \quad (i = 1, 2). \end{aligned} \quad (30)$$

Now put $\kappa_i = \tilde{\kappa}_i^*$ for $i = 1, 2$. Then $\kappa_1\kappa_2\kappa_3 = \kappa$, and (27) for $i = 1, 2$ follows from (28) and (30).

Case 2: where $p = 2$. According to Lemma 5.2 of [Sg], $f_i(U_i) = P_i$, say, is a subgroup of index at most 2 in A' for each i , and (30) holds for each $\tilde{\kappa}_i \in P_i$.

Subcase 2.1: $P_t \neq P_l$ for some pair t, l . Then $P_1 + P_2 + P_3 = A'$ so there exist $\tilde{\kappa}_i = f_i(u_i) \in P_i$ such that $\tilde{\kappa}_1 + \tilde{\kappa}_2 + \tilde{\kappa}_3 = \tilde{\kappa}$. Then both (26) and (27) hold with $\kappa_i = \tilde{\kappa}_i^*$, as in Case 1.

Subcase 2.2: $P_1 = P_2 = P_3 = P$ say, with $|A' : P| \leq 2$. According to the extra hypotheses in Case 2, there exists $a \in A^*$ such that $a^2 \equiv \kappa \pmod{P^*}$, and there exist $v_i \in A^{*(m)}$ such that $\phi_i(v_i) \equiv a \pmod{A^{*l}}$ (because A^*/A^{*l} is a perfect G^*/B^* -module and x_{i1}, \dots, x_{im} generate G^* modulo B^*). By the pigeonhole principle, there exist $t < l$ such that $a^{-1}\phi_t(v_t) \equiv a^{-1}\phi_l(v_l) \pmod{P^*}$; as A^{*l} has exponent 2 we then have

$$\phi_t(v_t)\phi_l(v_l) \equiv a^2 \equiv \kappa \pmod{P^*}.$$

Now put

$$\begin{aligned} \kappa_t &= \phi_t(v_t), \quad \kappa_l = \phi_l(v_l), \\ \kappa_j &= (\phi_t(v_t)\phi_l(v_l))^{-1} \kappa \end{aligned}$$

where $\{t, l, j\} = \{1, 2, 3\}$. Then

$$\kappa_j \in P^* = f_j(U_j)^* \subseteq \phi_j(A^{*(m)}),$$

and $\kappa_1\kappa_2\kappa_3 = \kappa$ since P^* is central.

To establish (27), it now suffices to show that for each i and each $v \in A^{*(m)}$, the fibre $\phi_i^{-1}(\phi_i(v))$ contains at least $|M|^{m-d-1}$ cosets of B^{*m} . Say $v = (z_1^*, \dots, z_m^*)$. Then for each $u \in (f_i^z)^{-1}(0)$ we have

$$\phi_i(z_1^*a_1^*, \dots, z_m^*a_m^*) = \phi_i(v)$$

whenever $(a_1B, \dots, a_mB) = u$, by (29). Our claim now follows from (30) with f_i^z in place of f_i and 0 for $\tilde{\kappa}_i$. ■

We turn now to the proof of Proposition 6.1. For $a \in A$ and $g \in G$ we shall write

$$\begin{aligned}\bar{g} &= gB \in G/B = \bar{G} \\ \bar{a} &= aB \in A/B = M \\ \tilde{a} &= aA' \in A/A' = \tilde{A}.\end{aligned}$$

Since $[B, A] = [A', G] = 1$, for $a, d \in A$ and $g \in G$ we may set

$$\begin{aligned}[\bar{a}, d] &= [a, d], \\ [\tilde{a}, g] &= [a, g], \\ a^{\bar{g}} &= a^g, [a, \bar{g}] = [a, g].\end{aligned}$$

Lemma 6.3 *Let $g_1, \dots, g_n \in \bar{G}$, $\varepsilon_1, \dots, \varepsilon_n \in \{1, -1\}$ satisfy*

$$\sum_{j=1}^n \varepsilon_j g_j = 0$$

in the group ring $R = \mathbb{Z}\bar{G}$. For $a \in A$ let

$$\psi(a) = \prod_{j=1}^n a^{\varepsilon_j g_j}.$$

Then there exist $h_i, k_i \in \{g_1, \dots, g_n\}$ ($i = 1, \dots, l$) such that for each $a \in A$

$$\psi(a) = \prod_{i=1}^l [a^{h_i}, a^{k_i}].$$

Proof. The hypothesis implies that $n = 2t$ is even and that the sequence $(\varepsilon_1 g_1, \dots, \varepsilon_n g_n) = \mathcal{S}$ is some re-arrangement of $(y_1, -y_1, \dots, y_t, -y_t) = \mathcal{S}'$ where each y_i is one of the g_j . Since A' is central in A , it follows that for each $a \in A$ we have

$$\psi(a) = \prod_{i=1}^t a^{y_i} a^{-y_i} \cdot \mathfrak{r}(a) = \mathfrak{r}(a)$$

where $\mathfrak{r}(a)$ is the product of certain factors of the form $[a^{\varepsilon_i g_i}, a^{\varepsilon_j g_j}]$, namely those for which $i < j$ while $\varepsilon_i g_i$ is moved to the right of $\varepsilon_j g_j$ when \mathcal{S} is re-arranged to \mathcal{S}' . The result follows since

$$[a^{\varepsilon_i g_i}, a^{\varepsilon_j g_j}] = \begin{cases} [a^{g_i}, a^{g_j}] & (\varepsilon_i \varepsilon_j = 1) \\ [a^{g_j}, a^{g_i}] & (\varepsilon_i \varepsilon_j = -1) \end{cases}.$$

■

Corollary 6.4 *In the notation of Lemma 6.3, if $aB = \bar{c}\mu$ with $\mu \in S$ then*

$$\psi(a) = \mu^2 \psi(c).$$

Now fix $x_1, \dots, x_m \in \overline{G}$ with $\langle x_1, \dots, x_m \rangle = \overline{G}$. Define mappings

$$f : R^m \rightarrow A$$

$$\mathbf{r} \mapsto \prod_{i=1}^m [\tilde{c}r_i, x_i],$$

and

$$B : R^m \times R^m \rightarrow A'$$

$$(\mathbf{r}, \mathbf{s}) \mapsto \prod_{i=1}^m [[\tilde{c}r_i, x_i], \tilde{c}s_i] \cdot \prod_{1 \leq i < j \leq m} [[\tilde{c}r_i, x_i], [\tilde{c}s_j, x_j]]$$

and

$$\Xi : R^m \rightarrow R(\overline{G} - 1)$$

$$\mathbf{r} \mapsto \sum_{i=1}^m r_i(x_i - 1).$$

(Here, $\mathbf{r} = (r_1, \dots, r_m)$ etc.)

The following observations are more or less immediate; note that identifying A' with $\wedge_R^2 M$ we can equally well write

$$B(\mathbf{r}, \mathbf{s}) = \sum_i \tilde{c}r_i(x_i - 1) \wedge \tilde{c}s_i + \sum_{i < j} \tilde{c}r_i(x_i - 1) \wedge \tilde{c}s_j(x_j - 1),$$

and that

$$\widetilde{f(\mathbf{r})} = \tilde{c}\Xi(\mathbf{r})$$

for each $\mathbf{r} \in R^m$.

Lemma 6.5 (i)

$$f(\mathbf{r} + \mathbf{s}) = f(\mathbf{r}) \cdot f(\mathbf{s}) \cdot B(\mathbf{r}, \mathbf{s}).$$

(ii) B is S -bilinear and $B(R^m, I^{(m)}) = B(I^{(m)}, R^m) = 0$.

(iii) Ξ is a left R -module epimorphism.

Now since $A = B\langle c^G \rangle$ and $A' \neq 1$ there exists $d \in A$ such that $[c, d] \neq 1$. We fix such a d .

Lemma 6.6 If $s \in S$ and $s(\overline{d} - 1) = 0$ in R then $s \in I$.

Proof. Say $s = \sum_{j=1}^n \varepsilon_j g_j$. Then $\sum_{j=1}^n \varepsilon_j g_j - \sum_{j=1}^n \varepsilon_j g_j \overline{d} = 0$, and Corollary 6.4 applies to the mapping ψ given by

$$\psi(a) = [\tilde{a}s, \overline{d}] \quad (a \in A).$$

Hence if $\tilde{a} = \tilde{c}\mu$, where $\mu \in S$, then

$$\psi(a) = \mu^2 \psi(c) = \mu^2 [\tilde{c}s, \overline{d}] = \mu^2 s[c, d].$$

On the other hand, we also have

$$\psi(a) = [\tilde{a}s, \bar{d}] = [\tilde{c}\mu s, \bar{d}] = \mu s[c, d].$$

Since $|k| > 2$ we may choose $\mu \in S$ so that $\mu - \mu^2 \not\equiv 0 \pmod{I}$, and deduce that $s[c, d] = 0$. The result follows since I is the annihilator of each non-zero element in A' . ■

Put

$$V = \Xi^{-1}(S(\bar{d} - 1)).$$

Thus V is a left S -submodule of R^m , and Ξ maps V onto $S(\bar{d} - 1)$. Moreover,

$$f(V) \subseteq A',$$

since if $\Xi(v) = s(\bar{d} - 1)$ then

$$\widetilde{f(v)} = \tilde{c}s(\bar{d} - 1) = 0.$$

In view of Lemma 6.6, there is a well-defined mapping

$$\alpha : V \rightarrow A'$$

such that

$$\alpha(v) = [\tilde{c}s, d] = s[c, d]$$

when $\Xi(v) = s(\bar{d} - 1)$, $s \in S$. Evidently α is a left S -module epimorphism. Define

$$\Phi : V \rightarrow A'$$

by

$$\Phi(v) = f(v) - \alpha(v).$$

Lemma 6.7 *For each $v \in V$ and $\mu \in S$ we have*

$$\Phi(\mu v) = \mu^2 \Phi(v).$$

Proof. Say $v = (r_1, \dots, r_m) \in V$ and $\Xi(v) = s(\bar{d} - 1)$ with $s \in S$. For $a \in A$ put

$$\psi(a) = \prod_{i=1}^m [\tilde{a}r_i, x_i] \cdot [\tilde{a}s, d]^{-1}.$$

Since $\sum r_i(x_i - 1) - s(\bar{d} - 1) = 0$, we may apply Corollary 6.4 to deduce that if $\tilde{a} = \tilde{c}\lambda$ where $\lambda \in S$ then $\psi(a) = \lambda^2 \psi(c)$. But

$$\prod_{i=1}^m [\tilde{a}r_i, x_i] = \prod_{i=1}^m [\tilde{c}\lambda r_i, x_i] = f(\lambda v)$$

and

$$[\tilde{a}s, d] = [\tilde{c}\lambda s, d] = \alpha(\lambda v),$$

so $\psi(a) = \Phi(\lambda v)$. Thus in particular $\Phi(v) = \psi(c)$ and $\Phi(\mu v) = \psi(a)$ where $\tilde{a} = \tilde{c}\mu$, and the lemma follows. ■

For $w \in R^m$ and $v \in V$ put

$$\alpha^w(v) = \alpha(v) + B(w, v).$$

Then $\alpha^w : V \rightarrow A'$ is a left S -module homomorphism, and Lemma 6.5 shows that

$$\begin{aligned} \Phi(v) + \alpha^w(v) &= f(v) + B(w, v) \\ &= f(w + v)f(w)^{-1} \end{aligned}$$

for each $v \in V$.

It follows from Lemma 6.5 that for $u, v \in V$,

$$\Phi(u + v) = \Phi(u) + \Phi(v) + B(u, v).$$

With Lemma 6.7 this implies that Φ factors through $V \rightarrow V/IV$, and that the mapping $\bar{\Phi} : V/IV \rightarrow A'$ induced by Φ is quadratic as a map of k -vector spaces; similarly, α^w factors through $V \rightarrow V/IV$ and induces a k -linear map $\bar{\alpha}^w : V/IV \rightarrow A'$.

Lemma 6.8 *Each of the maps Φ and α^w factors through $V \rightarrow (V + I^{(m)})/I^{(m)}$; and B factors through $R^m \times R^m \rightarrow R^m/I^{(m)} \times R^m/I^{(m)}$.*

Proof. The claim regarding B is immediate from Lemma 6.5(ii). For the rest, we separate two cases.

Case 1: $p \neq 2$. Let $v \in V$. Then

$$4\Phi(v) = \Phi(2v) = 2\Phi(v) + B(v, v)$$

so $\Phi(v) = \frac{1}{2}B(v, v)$ which depends only on the coset of v modulo $I^{(m)}$. In particular, $\Phi(I^{(m)} \cap V) = 0$, so $f(I^{(m)} \cap V) = \alpha(I^{(m)} \cap V)$ is a k -subspace of A' . But if $\mathbf{s} \in I^{(m)} \cap V$ then

$$f(\mathbf{s}) = \prod [\tilde{c}s_i, x_i] \in [B, G] \cap A'.$$

As $[B, G] \cap A' < A'$ and A' is a 1-dimensional k -space it follows that $\alpha(I^{(m)} \cap V) = f(I^{(m)} \cap V) = 0$. Since $B(w, I^{(m)}) = 0$ we also have $\alpha^w(I^{(m)} \cap V) = 0$.

Thus both Φ and α^w factor through $V \rightarrow (V + I^{(m)})/I^{(m)}$.

Case 2: $p = 2$. Consider the k -vector space $W = (I^{(m)} \cap V)/IV$, and write

$$\bar{f} = (\bar{\Phi} + \bar{\alpha})|_W : W \rightarrow A'.$$

Suppose that $\bar{f}(W) \neq \{0\}$. According to Lemma 5.2 of [Sg], $\bar{f}(W)$ is then an additive subgroup of index at most 2 in A' . However, $\bar{f}(W) = f(I^{(m)} \cap V) \subseteq$

$[B, G] \cap A'$, as observed above, so $|A' : [B, G] \cap A'| \leq 2$. This contradicts our original hypothesis; it follows that

$$f(I^{(m)} \cap V) = \bar{f}(W) = \{0\}$$

and hence that $\bar{\alpha}|_W = \bar{\Phi}|_W$. Since $\bar{\alpha}$ is linear, $\bar{\Phi}$ is quadratic and $|k| > 2$ it follows that $\bar{\alpha}|_W = \bar{\Phi}|_W = 0$.

Hence $\alpha(I^{(m)} \cap V) = \Phi(I^{(m)} \cap V) = 0$, and the proof is now completed as in Case 1. ■

Lemma 6.9 $\dim_k((V + I^{(m)})/I^{(m)}) \geq (m - d) \dim_k(M)$.

Proof. Put $h = \dim_k(M)$. Then R/I is generated as a left S -module by h elements, one of which may be taken to be 1_R ; as $I \subseteq S$ it follows that R is an h -generator left S -module. Since $R(\bar{G} - 1)$ is a d -generator left R -module, it is a dh -generator left S -module. As $V \supseteq \ker \Xi$ it follows that R^m/V is a dh -generator left S -module, and hence that $R^m/(V + I^{(m)})$ is a k -vector space of dimension at most dh . On the other hand, $R^m/I^{(m)} \cong M^{(m)}$ is a k -vector space of dimension mh . The lemma follows since

$$\dim_k((V + I^{(m)})/I^{(m)}) = \dim_k(R^m/I^{(m)}) - \dim_k(R^m/(V + I^{(m)})).$$

■

We can now complete the proof of Proposition 6.1. Fix $\mathbf{z} \in A^{(m)}$. Since $\tilde{A} = \tilde{B} + \tilde{c}R$ we can write $\tilde{z}_i = b_i + \tilde{c}t_i$ with $b_i \in B$ and $t_i \in R$, and we put

$$w = (t_1, \dots, t_m) \in R^m.$$

The map

$$(r_1, \dots, r_m) \mapsto (\bar{c}r_1, \dots, \bar{c}r_m)$$

induces a left S -module isomorphism $\theta : R^m/I^{(m)} \rightarrow M^{(m)}$. Put

$$U = \theta((V + I^{(m)})/I^{(m)}) \leq M^{(m)}.$$

Then U is a k -subspace of $M^{(m)}$ and Lemma 6.9 shows that $\dim_k U \geq (m - d) \dim_k M$. According to Lemma 6.8, the maps Φ and α^w induce maps $\tilde{\Phi}$ and $\tilde{\alpha}^w$ from $(V + I^{(m)})/I^{(m)}$ to A' . Then

$$\Phi_0 = \tilde{\Phi} \circ \theta^{-1} : U \rightarrow A'$$

is quadratic over k and

$$\alpha_0^w = \tilde{\alpha}^w \circ \theta^{-1} : U \rightarrow A'$$

is linear over k ; also α_0^0 is surjective; these all follow from the corresponding properties of $\bar{\Phi}$, $\bar{\alpha}^w : V/IV \rightarrow A'$ established above.

Let $u = \theta(\mathbf{r}) \in U$, and let $a_i \in A$ be such that $\tilde{a}_i = \tilde{c}r_i$. Then $(a_1B, \dots, a_mB) = u$. Now

$$\begin{aligned} \prod_{i=1}^m [z_i a_i, x_i] &= \prod_{i=1}^m [\tilde{b}_i + \tilde{c}(t_i + r_i), x_i] \\ &= \prod_{i=1}^m [\tilde{c}(t_i + r_i), x_i] \cdot \prod_{i=1}^m [\tilde{b}_i, x_i] \end{aligned} \quad (31)$$

since $[B, A] = 1$, and similarly

$$\prod_{i=1}^m [z_i, x_i] = \prod_{i=1}^m [\tilde{c}t_i, x_i] \cdot \prod_{i=1}^m [\tilde{b}_i, x_i]. \quad (32)$$

On the other hand,

$$\begin{aligned} \Phi_0(u) + \alpha_0^w(u) &= \Phi(\mathbf{r}) + \alpha^w(\mathbf{r}) \\ &= f(\mathbf{t} + \mathbf{r})f(\mathbf{t})^{-1} \\ &= \prod_{i=1}^m [\tilde{c}(t_i + r_i), x_i] \cdot \left(\prod_{i=1}^m [\tilde{c}t_i, x_i] \right)^{-1}. \end{aligned}$$

With (31) and (32) this shows that

$$\Phi_0(u) + \alpha_0^w(u) = \left(\prod_{i=1}^m [z_i a_i, x_i] \right) \cdot \left(\prod_{i=1}^m [z_i, x_i] \right)^{-1}.$$

This is precisely claim (ii) of Proposition 6.1, if we write Φ for Φ_0 and $\alpha^{\mathbf{z}}$ for α_0^w . In the special case $\mathbf{z} = (1, \dots, 1)$ we can take $\mathbf{t} = \mathbf{0}$ to ensure that α_0^w is surjective. This completes the proof.

7 The second inequality, soluble case

We are now ready to establish one of the main steps in the proof of the Key Theorem, concerning the case where N is a soluble quasi-minimal normal subgroup of G . The following notation and hypotheses are in force throughout this section.

G is a finite d -generator group, N is a soluble quasi-minimal normal subgroup of G , $Z = Z_N$ is the maximal normal subgroup of G properly contained in N , and we write $M = N/Z$. We assume in addition that $[Z, G] = 1$.

Recall (Lemma 4.2) that M is a simple $\mathbb{F}_p G$ -module for some prime p and that $[N, G] = N$,

$$\begin{aligned} N^p &= 1 \quad \text{if } p \neq 2 \\ N^2 &= N' \quad \text{and } N'^2 = 1 \quad \text{if } p = 2. \end{aligned}$$

Note that if $N' \neq 1$ then N/Z cannot be cyclic so $|M| \geq p^2$.

Set

$$K = \begin{cases} N & \text{if } N' = 1 \\ N' & \text{if } N' > 1 \end{cases}.$$

For $i = 1, 2, 3$ let x_{i1}, \dots, x_{im} satisfy $K \langle x_{i1}, \dots, x_{im} \rangle = G$. Define

$$\begin{aligned} \phi_i : N^{(m)} &\rightarrow N \\ (a_1, \dots, a_m) &\mapsto \prod_{j=1}^m [a_j, x_{ij}]. \end{aligned}$$

Proposition 7.1 *Let $\kappa \in K$. Then there exist $\kappa_1, \kappa_2, \kappa_3 \in N$ such that*

$$\kappa_1 \kappa_2 \kappa_3 = \kappa$$

and for $i = 1, 2, 3$

$$|\phi_i^{-1}(\kappa_i)| \geq |N|^m |M|^{-d-1}. \quad (33)$$

Proof. Note that (33) holds if (and only if) $\phi_i^{-1}(\kappa_i)$ contains at least $|M|^{m-d-1}$ cosets of $Z^{(m)}$. We separate cases.

Case 1: where $N = K$ is abelian. Write N additively, and suppose that $G = \langle g_1, \dots, g_d \rangle$. The mapping $(a_1, \dots, a_d) \mapsto \sum a_i(g_i - 1)$ induces an epimorphism from $M^{(d)}$ to $[N, G] = N$, so $|N| \leq |M|^d$. Similarly, ϕ_i induces an epimorphism $\bar{\phi}_i : M^{(m)} \rightarrow N$. Take $\kappa_1 = \kappa$ and $\kappa_2 = \kappa_3 = 1$. Now $\phi_i^{-1}(\kappa_i)$ consists of $|\bar{\phi}_i^{-1}(\kappa_i)|$ cosets of $Z^{(m)}$, and the result follows since

$$|\bar{\phi}_i^{-1}(\kappa_i)| = |\ker \bar{\phi}_i| = |M^m| / |N| \geq |M|^{m-d}.$$

Case 2: where $|N'| = 2$, $K = N'$. Let $\bar{\phi}_i : M^{(m)} \rightarrow N$ and $\tilde{\phi}_i : M^{(m)} \rightarrow N/N'$ denote the maps naturally induced by ϕ_i . As above, each $\tilde{\phi}_i$ is an epimorphism, and each fibre of $\tilde{\phi}_i$ has size at least $|M|^{m-d}$. There exists $c \in N$ with $c^2 \neq 1$, and then $N' = \{1, c^2\}$. For each i we now have

$$\bar{\phi}_i^{-1}(c) \cup \bar{\phi}_i^{-1}(c^3) = \tilde{\phi}_i^{-1}(cN')$$

so $|\bar{\phi}_i^{-1}(c^{\varepsilon(i)})| \geq \frac{1}{2} |M|^{m-d} \geq |M|^{m-d-1}$ where $\varepsilon(i)$ is 1 or 3. One of these two values must occur at least twice as i ranges over $\{1, 2, 3\}$; say $\varepsilon(s) = \varepsilon(t) = \varepsilon$. Now if $\kappa = c^2$ put $\kappa_s = \kappa_t = c^\varepsilon$, $\kappa_u = 1$ where $\{1, 2, 3\} = \{s, t, u\}$; if $\kappa = 1$ put $\kappa_1 = \kappa_2 = \kappa_3 = 1$. In either case we then have $\kappa_1 \kappa_2 \kappa_3 = \kappa$ (since $c^6 = c^2$).

Now $\phi_i^{-1}(\kappa_i)$ is the union of $|\bar{\phi}_i^{-1}(\kappa_i)|$ cosets of $Z^{(m)}$; so to complete the proof in this case it remains to show that $|\bar{\phi}_i^{-1}(1)| \geq |M|^{m-d-1}$. Put $V = \bar{\phi}_i^{-1}(N') =$

$\ker \tilde{\phi}_i$, so $|V| \geq |M|^{m-d}$. We claim that $\bar{\phi}_i|_V : V \rightarrow N'$ is a quadratic form over \mathbb{F}_2 , if N' is identified with \mathbb{F}_2 . To see this, define $B : V \times V \rightarrow N'$ by

$$B(\mathbf{u}, \mathbf{v}) = \bar{\phi}_i(\mathbf{u} + \mathbf{v}) - \bar{\phi}_i(\mathbf{u}) - \bar{\phi}_i(\mathbf{v});$$

one readily verifies that if $\mathbf{u} = (u_1 Z, \dots, u_m Z)$, $\mathbf{v} = (v_1 Z, \dots, v_m Z)$ then

$$B(\mathbf{u}, \mathbf{v}) = \sum_{j=1}^m [[u_j, x_{ij}], v_j] + \sum_{j < l} [[u_j, x_{ij}], [v_l, x_{il}]],$$

and hence that B is bilinear as a map of \mathbb{F}_2 -spaces. This establishes the claim, which then implies that each fibre of $\bar{\phi}_i|_V$ has size at least

$$2^{\dim_{\mathbb{F}_2}(V)-2} = \frac{1}{4} |V| \geq |M|^{m-d-1}$$

(cf. Lemma 5.2 of [Sg]). The result follows.

Case 3: where $|N'| > 2$, $K = N'$. Let F be a free group and $\pi : F \rightarrow G$ an epimorphism. Set $A = \pi^{-1}(N)$ and $B = \pi^{-1}(Z)$. Then A is free, and it is well known that the mapping $(a, b) \mapsto [a, b]$ induces an isomorphism

$$\theta_1 : A/A' \wedge A/A' \rightarrow A'/[A', A].$$

Write $M_1 = A/B$. Noting that $A'A^p \leq B$, one verifies easily that θ_1 induces an isomorphism

$$\theta_2 : M_1 \wedge M_1 \rightarrow A'/[B, A].$$

The group F/B acts by conjugation on $A/[B, A]$; and θ_2 becomes an isomorphism of $R = \mathbb{Z}(F/B)$ -modules when F/B is made to act diagonally on $A/B \wedge A/B$, so θ_2 induces an isomorphism

$$\theta_3 : \wedge_R^2 M_1 = \frac{M_1 \wedge M_1}{[M_1 \wedge M_1, F]} \rightarrow \frac{A'}{[A', F][B, A]}.$$

Now let $\bar{\cdot} : F \rightarrow F/[A', F][B, A]$ denote the quotient map. Since $[N', G][Z, N] \leq [Z, G] = 1$, the map π induces an epimorphism $\ast : \bar{F} \rightarrow G$. Evidently

$$\begin{aligned} [\bar{B}, \bar{A}] &= [\bar{A}', \bar{F}] = 1, \\ \ker(\ast) &\leq \bar{B}, \quad \bar{A}^\ast = N, \quad \bar{B}^\ast = Z, \end{aligned}$$

and

$$|\bar{A}' : \bar{A}' \cap [\bar{B}, \bar{F}]| \geq |N'| > 2.$$

Thus all the hypotheses of Section 6 are satisfied if we take \bar{F} for G , \bar{A} for A and \bar{B} for B ; Proposition 7.1 thus reduces in the present case to an application of Proposition 6.2, with G taking the role of G^\ast .

This completes the proof. ■

8 Word combinatorics

In the next three sections we examine the solution of equations in a direct product of quasisimple groups. This preparatory section is devoted to some observations on the shape of abstract group words, generalizing Lemma 1 of [N2]: these will help us to keep track of the equations as the unknowns are successively eliminated.

The material here is rather abstract, and won't make much sense until it is applied. However, it seems inevitable, given the nature of our main theorems, that at some stage we will have to get to grips with the detailed rewriting of words in a group; by separating off in this section some of the most technical steps, we hope to make the complicated arguments of the later sections a little less opaque.

Let Γ be a group and Y a non-empty set. The *free Γ -group on Y* is the free group on the alphabet $Y^\Gamma = \{y^g \mid y \in Y, g \in \Gamma\}$, on which Γ acts by permuting the basis in the obvious way. We denote it by

$$F_\Gamma(Y);$$

it may be identified with the normal closure of the free group $F(Y)$ in the free product $F(Y) * \Gamma$.

A subset Z of $F_\Gamma(Y)$ will be called *independent* if every map from Z into an arbitrary Γ -group S can be extended to a Γ -equivariant homomorphism from F to S (thus for example every subset of Y is independent). The following 'invariance' and 'exchange' principles are more or less self-evident: (i) if Z is independent and $g(y) \in \Gamma$ for each $y \in Z$ then $\{y^{g(y)} \mid y \in Z\}$ is independent; (ii) if $Z \cup \{x\}$ is independent and $P, Q \in \langle Z^\Gamma \rangle$ then $Z \cup \{PxQ\}$ is independent. A family of elements $\{z_1, z_2, \dots\}$ is called independent if its terms are all distinct and form an independent set.

As a matter of notation, we will usually write y for y^1 and y^{-g} in place of $(y^g)^{-1}$ ($y \in Y, g \in \Gamma$).

Now we fix two disjoint sets, a set X of *variables* and a set P of *parameters*, and consider the free Γ -group

$$F = F_\Gamma(X \cup P).$$

Let \mathcal{M} denote the the free monoid on the set $\{y^{\pm g} \mid y \in X \cup P, g \in \Gamma\}$; this is the set of 'unreduced' group words on the alphabet $X^\Gamma \cup P^\Gamma$. Let $W \subseteq \mathcal{M}$ denote the free monoid on $X \cup X^{-1}$. There is a natural map $\overline{} : \mathcal{M} \rightarrow F$ (evaluation), and we define a mapping $\hat{} : \mathcal{M} \rightarrow W$ as follows: for $U \in \mathcal{M}$, let $\hat{U} \in W$ denote the word obtained from U by deleting all terms belonging to $P^\Gamma \cup P^{-\Gamma}$ and replacing each term $x^{\pm g}$ with $x^{\pm 1}$ ($x \in X, g \in \Gamma$).

For $U, V \in \mathcal{M}$ we write

$$U =_F V \iff \overline{U} = \overline{V}$$

(the notation $U = V$ for $U, V \in \mathcal{M}$ will always mean that U and V are identical as words).

We write

$$|x| = \begin{cases} x & (x \in X) \\ x^{-1} & (x \in X^{-1}) \end{cases},$$

and for $w \in W$ put

$$\sup(w) = \{|x| \mid x \text{ occurs in } w\}.$$

We call $w \in W$ *balanced* if each element of $\sup(w) \cup \sup(w)^{-1}$ occurs exactly once in w .

Lemma 8.1 *Suppose that $w \in W$ is balanced and $w \neq_F 1$. Then*

$$w = Ax^{-1}By^{-1}CxDyE \quad (34)$$

for some $x, y \in X \cup X^{-1}$ with $|x| \neq |y|$ and $\{|x|, |y|\} \cap \sup(ABCDE) = \emptyset$.

Proof. The hypotheses imply that $w = u_1y^{-1}vyu_2$ where $y \in X \cup X^{-1}$ and $v \neq \emptyset$. Choose such an expression with v as short as possible. Say x occurs in v where $x \in X \cup X^{-1}$. Then x^{-1} must occur in u_1 or in u_2 ; in the first case we have (34), in the second case we get (34) on replacing x by x^{-1} and then interchanging x and y . The final claim is clear since w is balanced. ■

Proposition 8.2 *Let $V \in \mathcal{M}$. Suppose that \widehat{V} is balanced and $\widehat{V} \neq_F 1$. Then*

$$V =_F T_{a,b}(\xi, \eta) \cdot V_1$$

for some $a, b \in \Gamma$ and $\xi, \eta, V_1 \in \mathcal{M}$ such that (i) the family $\{\bar{\xi}, \bar{\eta}\} \cup \sup(\widehat{V}_1) \cup P$ is independent, and (ii) ignoring exponents from Γ , each term from $P \cup P^{-1}$ occurs with the same multiplicity in V_1 as it has in V .

Recall that

$$T_{a,b}(\xi, \eta) = \xi^{-1}\eta^{-1}\xi^a\eta^b.$$

Proof. Lemma 8.1 shows that

$$\widehat{V} = Ax^{-1}By^{-1}CxDyE,$$

say, for suitable words A, B etc. in W and $x, y \in X \cup X^{-1}$ with $|y| \neq |x|$, $\{|x|, |y|\} \cap \sup(ABCDE) = \emptyset$. It follows that

$$V = A'x^{-e}B'y^{-f}C'x^{ea}D'y^{fb}E'$$

where $a, b, e, f \in \Gamma$ and $A', B', \dots \in \mathcal{M}$ satisfy $\widehat{A'} = A$, $\widehat{B'} = B$ etc. Now put

$$\begin{aligned} U_1 &= A'^{ab^{-1}}D'^{b^{-1}}, \quad U_2 = U_1^{a^{-1}}C'^{a^{-1}}, \\ \xi &= U_2x^eA'^{-1}, \\ \eta &= U_1y^fB'^{-1}U_2^{-1}. \end{aligned}$$

A direct calculation shows that

$$V =_F T_{a,b}(\xi, \eta) \cdot V_1$$

where

$$V_1 = A'^{ab^{-1}a^{-1}b} D'^{b^{-1}a^{-1}b} C'^{a^{-1}b} B'^b E'.$$

Note that $\widehat{V_1} = ADCBE$. The claim (i) follows from the invariance and exchange principles, and the claim (ii) is clear. ■

For the next proposition we need some further notation. Fix a mapping $\chi : X \rightarrow \{1, \dots, m\}$, and for each $x \in X$ define $\chi(x^{-1}) = -\chi(x)$. We call $\chi(x)$ the *colour* of x . For $w = y_1 y_2 \dots y_k \in W$ (with each $y_i \in X \cup X^{-1}$) define $\chi(w)$ to be the sequence

$$\chi(w) = (\chi(y_1), \dots, \chi(y_k)).$$

A new sequence $\tau(w)$, the *colour type* of w , is now defined as follows: first, wherever a segment consisting of consecutive equal negative terms occurs in $\chi(w)$, delete all but one of them (so a maximal segment $(-r, -r, \dots, -r)$ is contracted to $-r$); then replace each term by its absolute value. For example $(1, 1, -2, 2, -2, -2, -3) \mapsto (1, 1, -2, 2, -2, -3) \mapsto (1, 1, 2, 2, 2, 3)$.

For sequences S and T we write

$$S \leq T$$

to indicate that S is a subsequence of T . Put

$$L_n = (1, 2, \dots, m, 1, 2, \dots, m, \dots, 1, 2, \dots, m)$$

with n repetitions of $(1, 2, \dots, m)$.

Lemma 8.3 *Let $w \in W$ be balanced. Put $Y = \text{sup}(w)$, and suppose that $\tau(w) \leq L_n$, where $1 \leq n < |Y|$. Then there exist $x, y \in Y \cup Y^{-1}$ with $|y| \neq |x|$ such that*

$$w = Ax^{-1}By^{-1}Cx Dy E \tag{35}$$

where $w_0 = ADCBE$ is balanced and $\tau(w_0) \leq L_n$.

Proof. We claim that $w \neq_F 1$. The proof is by induction on n . Suppose that $w =_F 1$. Then $w = uxx^{-1}v$ where $x \in X \cup X^{-1}$. Suppose that $x \in X$ and x has colour i . Then $\tau(w) = (\tau(u), i, i, S)$ where $\tau(v)$ is either S or (i, S) . In either case, $\tau(uv) \leq (\tau(u), i, S) \leq L_{n-1}$. One sees similarly that $\tau(uv) \leq L_{n-1}$ if $x \in X^{-1}$. As uv is balanced and $|\text{sup}(uv)| = |\text{sup}(w)| - 1 > n - 1$ the inductive hypothesis gives $uv \neq_F 1$, a contradiction.

Applying Lemma 8.1 we obtain the expression (35).

It is clear that w_0 is again balanced. To establish the final claim, suppose for example that $\chi(x) = i > 0$ and $\chi(y) = j > 0$. Let A_0, B_0, C_0 be the words obtained from A, B, C respectively by removing all terms coloured $-i$ from the

end of A and the beginning of B , and all terms coloured $-j$ from the end of B and the beginning of C . Unless $B_0 = \emptyset$ and $i = j$ we then have

$$\begin{aligned}\tau(w) &= (\tau(A_0), i, \tau(B_0), j, \tau(C_0), i, \tau(D), j, \tau(E)), \\ \tau(w_0) &\leq (\tau(A_0), i, \tau(D), j, \tau(C_0), i, \tau(B_0), j, \tau(E)).\end{aligned}$$

It is easy to see that if the first sequence is a subsequence of L_n , then so is the second. The other cases are dealt with similarly. ■

Proposition 8.4 *Let $V \in \mathcal{M}$ and $k \in \mathbb{N}$, and put $w = \widehat{V}$. Suppose (a) w is balanced and (b) $\tau(w) \leq L_n$ for some $n \geq 1$ with*

$$|\sup(w)| \geq n + 2k.$$

Then there exist $V_k, \xi_1, \eta_1, \dots, \xi_k, \eta_k \in \mathcal{M}$ such that $|\sup(\widehat{V_k})| = |\sup(w)| - 2k$ and

- (i) $\{\overline{\xi_1}, \overline{\eta_1}, \dots, \overline{\xi_k}, \overline{\eta_k}\} \cup \sup(\widehat{V_k}) \cup P$ *is an independent family;*
- (ii)

$$V =_F T_{a_1, b_1}(\xi_1, \eta_1) \cdot \dots \cdot T_{a_k, b_k}(\xi_k, \eta_k) \cdot V_k$$

for some $a_i, b_i \in \Gamma$.

Proof. Put $Y = \sup(w)$. Lemma 8.3 shows that

$$w = Ax^{-1}By^{-1}CxDyE,$$

say, for suitable words A, B, \dots in W and $x, y \in Y \cup Y^{-1}$ with $|y| \neq |x|$. We may now define $\xi_1 = \xi, \eta_1 = \eta, a_1 = a, b_1 = b$ and V_1 as in the proof of Proposition 8.2 above, to obtain

$$V =_F T_{a_1, b_1}(\xi_1, \eta_1) \cdot V_1,$$

where $\widehat{V_1} = ADCBE$ and the family $\{\overline{\xi_1}, \overline{\eta_1}\} \cup \sup(\widehat{V_1}) \cup P$ is independent. Evidently

$$|\sup(\widehat{V_1})| = |\sup(w)| - 2.$$

If $k = 1$ we are done.

Suppose that $k > 1$. Put $w_1 = \widehat{V_1}$. According to Lemma 8.3 the word w_1 is balanced and satisfies $\tau(w_1) \leq L_n$. Also

$$\begin{aligned}|\sup(w_1)| &= |\sup(w) \setminus \{|x|, |y|\}| \\ &\geq n + 2(k - 1).\end{aligned}$$

Arguing by induction on k , we may therefore suppose that

$$V_1 =_F T_{a_2, b_2}(\xi_2, \eta_2) \cdot \dots \cdot T_{a_k, b_k}(\xi_k, \eta_k) \cdot V_k$$

is of the required form, and the result follows. ■

9 Equations in semisimple groups, 1: the second inequality

Let N be a quasi-semisimple group with centre Z and let g_1, \dots, g_m be automorphisms of N . We assume that N/Z is the direct product of $n \geq 2$ simple groups, and that the group generated by g_1, \dots, g_m permutes these transitively.

For each i let $c(g_i)$ denote the number of cycles in this permutation representation of g_i . We shall establish the following, where $D \in \mathbb{N}$ is the absolute constant appearing in Theorem 1.9:

Proposition 9.1 *Suppose that*

$$\sum_{i=1}^m c(g_i) \leq (m-2)n - 2D. \quad (36)$$

Then for each $\kappa \in N$ the number of solutions $\mathbf{u} = (u_1, \dots, u_m) \in N^{(m)}$ to the equation

$$\kappa = [\mathbf{u}, \mathbf{g}] := \prod_{i=1}^m [u_i, g_i] \quad (37)$$

is at least $|N|^m |N/Z|^{-4D}$.

Before proving this, let us deduce the version used in §4 for the proof of the Key Theorem:

Proposition 9.2 *Let G be a finite group and N a quasi-semisimple quasi-minimal normal subgroup of G , such that N/Z_N is not simple. Suppose that $G = \langle y_1, \dots, y_m \rangle N$, and that the m -tuple \mathbf{y} has the (k, ε) fixed-point property where $k\varepsilon \geq 2D + 4$. Define $\phi : N^{(m)} \rightarrow N$ by*

$$\phi(\mathbf{a}) = \prod_{i=1}^m [a_i, y_i].$$

Then for each $\kappa \in N$ we have

$$|\phi^{-1}(\kappa)| \geq |N|^m |N/Z_N|^{-4D}.$$

The various terms used in this statement were introduced in Section 4. Rather than repeating the definitions wholesale, we recall those consequences that are relevant here: these may be taken as the hypotheses for Proposition 9.2.

- The normal subgroup N satisfies $N = [N, N] > Z_N = Z(N)$, and $N/Z_N = T_1 \times \dots \times T_n$ where $n \geq 2$ and the T_i are isomorphic simple groups;
- The conjugation action of G permutes the set $\{T_1, \dots, T_n\}$ transitively, and at least k of the y_j move at least εn of the T_i .

Proof. Apply Proposition 9.1, taking g_i to be the image of y_i in $\text{Aut}(N)$. It is only necessary to verify the condition (36). Let $|\text{fix}(g_j)|$ denote the number of fixed points of g_j in the set $\{T_1, \dots, T_n\}$. Then

$$n \geq |\text{fix}(g_j)| + 2(c(g_j) - |\text{fix}(g_j)|),$$

and for at least k values of j we have $|\text{fix}(g_j)| \leq n - \varepsilon n$. Therefore

$$2 \sum_{i=1}^m c(g_i) \leq \sum_{i=1}^m (n + |\text{fix}(g_i)|) \leq k(2n - \varepsilon n) + (m - k) \cdot 2n$$

and (36) follows since $k\varepsilon \geq 2D + 4$ and $n \geq 2$. ■

We proceed to prove Proposition 9.1, and from now on write $G = \langle g_1, \dots, g_m \rangle$. The universal cover of N is a direct product $\tilde{N} = S_1 \times \dots \times S_n$ where each S_i is a quasisimple group of universal type and $n \geq 2$. The action of G on N lifts to an action on \tilde{N} , and G permutes $\{S_1, \dots, S_n\}$ the same way it permutes the simple factors of N/Z .

Now $N = \tilde{N}/A$ for some $A \leq \tilde{Z} = Z(\tilde{N})$. If the proposition holds with \tilde{N} in place of N , and $\tilde{\kappa}$ is a preimage of κ , then $[\tilde{\mathbf{u}}, \mathbf{g}] = \tilde{\kappa}$ holds for at least $|\tilde{N}|^m |\tilde{N}/\tilde{Z}|^{-4D}$ values of $\tilde{\mathbf{u}} \in \tilde{N}^{(m)}$. These project to at least

$$\frac{|\tilde{N}|^m |\tilde{N}/\tilde{Z}|^{-4D}}{|A|^m} = |N|^m |N/Z|^{-4D}$$

solutions \mathbf{u} of (37) in $N^{(m)}$. Thus we may, and shall, assume henceforth that $N = \tilde{N}$.

By way of notation we shall write

$$S_i^{g^{-1}} = S_{g_i} \quad (g \in G, 1 \leq i \leq n)$$

(so $i \mapsto g_i$ gives the left action of G on $\{1, \dots, n\}$ corresponding to its right action on $\{S_1, \dots, S_n\}$). Since the action is transitive, the groups S_i are all isomorphic; we fix an identification of each S_i with a fixed quasisimple group S . Thus elements of N will be written in the form

$$x = (x(1), x(2), \dots, x(n))$$

with each $x(i) \in S$, and the action of G takes the form

$$x^g = (x^{(g1)g(1)}, x^{(g2)g(2)}, \dots, x^{(gn)g(n)});$$

here $g(i) \in \text{Aut}(S)$ is induced by $g|_{S_{(g_i)}} : S_{(g_i)} \rightarrow S_i$ (when each S_j is identified with S).

For a subset Δ of $\{1, \dots, n\}$,

$$\pi_\Delta : N \rightarrow \prod_{i \in \Delta} S_i$$

will denote the natural projection.

We are going to think of (37) as the equation

$$\kappa = x_1 x_2 \dots x_m, \quad (38)$$

to be solved for $x_1, \dots, x_n \in N$ subject to the conditions

$$x_i \in [N, g_i] \quad (39)$$

for each i . The equation (38) is equivalent to the system of equations $\mathcal{F} = (F_1, \dots, F_n)$:

$$\kappa(s) = x_1(s) x_2(s) \dots x_m(s). \quad (F_s)$$

To analyse the condition (39), let Ω_i denote the set of orbits of g_i on the set $\{1, \dots, n\}$. Then $x \in [N, g_i]$ if and only if $\pi_\Delta(x) \in [\pi_\Delta(N), g_i]$ for each orbit $\Delta \in \Omega_i$. For each such Δ let k_Δ be the first member of Δ and put $n(\Delta) = |\Delta|$. Then $g_i^{n(\Delta)}$ maps S_{k_Δ} to itself, inducing the automorphism $\beta_i(\Delta) = g^{n(\Delta)}(k_\Delta)$ of S .

We claim that $\pi_\Delta(x_i) \in [\pi_\Delta(N), g_i]$ if and only if there exists $u_i(\Delta) \in S$ such that

$$\prod_{j=0}^{n(\Delta)-1} x_i(g_i^j k_\Delta)^{g_i^j(k_\Delta)} = u_i(\Delta)^{-1} u_i(\Delta)^{\beta_i(\Delta)}. \quad (H_{i,\Delta})$$

Indeed, dropping the subscript i for the moment and putting $k = k_\Delta$, $n = n(\Delta)$, if $\pi_\Delta(x) = [\pi_\Delta(u), g_i]$ then

$$\begin{aligned} x(k) &= u(k)^{-1} u(gk)^{g(k)} \\ x(gk) &= u(gk)^{-1} u(g^2 k)^{g(gk)} \\ &\vdots \\ x(g^{n-1} k) &= u(g^{n-1} k)^{-1} u(g^n k)^{g(g^{n-1} k)} \end{aligned} \quad (40)$$

and $(H_{i,\Delta})$ follows with $u_i(\Delta) = u(k)$ (note that

$$g(g^{r-1} k) \dots g(gk) g(k) = g^r(k)$$

for each r). Conversely, if $(H_{i,\Delta})$ holds then putting $u(k) = u_i(\Delta)$ we can solve (40) for $u(gk)$, $u(g^2 k)$, \dots , $u(g^{n-1} k)$ in turn and so determine $\pi_\Delta(u)$ with $\pi_\Delta(x) = [\pi_\Delta(u), g_i] = \pi_\Delta(x)$. This establishes the claim.

Put

$$\begin{aligned} X &= \{x_i(s) \mid 1 \leq i \leq m, 1 \leq s \leq n\}, \\ \mathcal{U} &= \{u_i(\Delta) \mid 1 \leq i \leq m, \Delta \in \Omega_i\}, \\ \mathcal{K} &= \{\kappa(1), \kappa(2), \dots, \kappa(n)\}, \\ P &= \mathcal{U} \cup \mathcal{K}. \end{aligned}$$

Note that

$$|\mathcal{U}| = \sum_{i=1}^m |\Omega_i| = \sum_{i=1}^m c(g_i).$$

We start by considering these as sets of abstract symbols, and call P the set of *parameters* and X the set of *variables*. Each term $x_i(s)$ is assigned the *colour* i . We shall apply the results of Section 8, taking $\Gamma = \text{Aut}(S)$ and $F = F_\Gamma(X \cup P)$.

We are going to reduce the system \mathcal{F} subject to the conditions (39) to a single equation. First of all, for each i and each $\Delta \in \Omega_i$ we solve equation $(H_{i,\Delta})$ for $x_i(k_\Delta)$ and substitute the resulting expression in equation (F_{k_Δ}) . That is, replace $x_i(k_\Delta)$ by

$$u_i(\Delta)^{-1} u_i(\Delta)^{\beta_i(\Delta)} \cdot \left(\prod_{j=1}^{n(\Delta)-1} x_i(g_i^j k_\Delta)^{g_i^j(k_\Delta)} \right)^{-1}.$$

At this stage, the conditions (39) and all the variables $x_i(k_\Delta)$ have been eliminated, at the cost of introducing some parameters from \mathcal{U} . Call the resulting system of equations $\mathcal{F}' = (F'_1, \dots, F'_n)$, and let U_s be the word on $X^\Gamma \cup P^\Gamma$ on the right-hand side of F'_s .

Together, the words $\widehat{U}_1, \dots, \widehat{U}_n$ contain the variables

$$x_i(s), x_i(s)^{-1} \quad (s \neq k_\Delta \text{ for } \Delta \in \Omega_i),$$

that is,

$$mn - \sum_{i=1}^m |\Omega_i| = mn - \sum_{i=1}^m c(g_i)$$

matching pairs x, x^{-1} .

Recalling the definition of *colour type* from the previous section, observe also that the colour type of each \widehat{U}_s satisfies

$$\tau(\widehat{U}_s) \leq L_1 = (1, \dots, m).$$

Next, we successively reduce the number of equations by a process of substitution of variables. Suppose that $x \in X \cup X^{-1}$ occurs in U_1 but x^{-1} does not; then x^{-1} appears in U_l for some $l \neq 1$. Solve F'_l for x and substitute the resulting expression in U_1 . We call this a substitution ($l \rightarrow 1$). Each such operation reduces by one both the number of equations in \mathcal{F}' and the total number of variables. We claim now that it is possible to apply $n - 1$ substitutions and thus reach an equivalent system consisting of the single equation

$$\kappa(1) = U, \tag{41}$$

where U is a certain word on $X^\Gamma \cup P^\Gamma$.

To establish the claim, let us call two equations F'_s, F'_t *linked* if they share a variable from X (which then must appear with positive exponent in one of them and negative exponent in the other), and let \mathcal{R} be the equivalence relation on

\mathcal{F}' generated by the linked pairs. Now F'_s and F'_t are linked precisely when s and t lie in the same orbit of g_i for some i . As $G = \langle g_1, \dots, g_m \rangle$ acts transitively on $\{1, \dots, n\}$ it follows that \mathcal{F}' consists of one equivalence class under \mathcal{R} . Now a substitution $(l \rightarrow 1)$ eliminates only the variable used to link F'_l with F'_1 , and simultaneously eliminates the equation F'_l ; so the resulting system of equations \mathcal{F}'' still consists of a single \mathcal{R} -equivalence class. If $|\mathcal{F}''| > 1$ there exists $l' \neq 1$ such that the new equation F''_1 is linked to $F'_{l'} \in \mathcal{F}''$, and we can perform a substitution $(l' \rightarrow 1)$. Evidently the process may be repeated as long as more than one equation remains in the system, and the claim is now clear.

Since each substitution eliminates precisely one pair x, x^{-1} ($x \in X$), the word \widehat{U} is balanced and

$$|\sup(\widehat{U})| = mn - \sum_{i=1}^m c(g_i) - (n-1) \geq n + 2D + 1.$$

Moreover, we claim that $\tau(\widehat{U}) \leq L_n$. To see this, suppose that after $f-1$ substitutions U_1 has been transformed into V_f where $\tau(\widehat{V}_f) \leq L_f$. The next substitution $(l \rightarrow 1)$ has one of the following effects on \widehat{V}_f :

$$\begin{aligned} \widehat{V}_f = A_0 \alpha x^{-1} \beta B_0 &\mapsto \widehat{V_{f+1}} = A_0 \alpha \cdot DC \cdot \beta B_0 \\ \widehat{V}_f = A_0 \alpha x \beta B_0 &\mapsto \widehat{V_{f+1}} = A_0 \alpha \cdot DC \cdot \beta B_0 \end{aligned}$$

where $\tau(\alpha x^{-1} \beta) \leq L_1$ and $\widehat{U}_l = CxD$ in the first case, $\tau(\alpha x \beta) \leq L_1$ and $\widehat{U}_l = Cx^{-1}D$ in the second case. Say $\chi(x) = i$. Since $\tau(\widehat{U}_l) \leq L_1$, in the first case we have

$$\begin{aligned} \tau(C) &\leq (1, \dots, i-1), & \tau(D) &\leq (i+1, \dots, m), \\ \tau(\alpha) &\leq (1, \dots, i), & \tau(\beta) &\leq (i, \dots, m), \end{aligned}$$

while in the second case

$$\begin{aligned} \tau(C) &\leq (1, \dots, i), & \tau(D) &\leq (i, \dots, m), \\ \tau(\alpha) &\leq (1, \dots, i-1), & \tau(\beta) &\leq (i+1, \dots, m). \end{aligned}$$

In either case, $\tau(\alpha \cdot DC \cdot \beta) \leq L_2$. Therefore $\tau(\widehat{V_{f+1}}) \leq L_{f+1}$, and the claim follows by induction.

We may now apply Proposition 8.4, which shows that

$$U =_F T_{a_1, b_1}(x_1, y_1) \cdot \dots \cdot T_{a_D, b_D}(x_D, y_D) \cdot U_0$$

where $a_i, b_i \in \Gamma$ and

$$\{\overline{x_1}, \overline{y_1}, \dots, \overline{x_D}, \overline{y_D}\} \cup \sup(\widehat{U_0}) \cup \mathcal{U} \cup \mathcal{K} \quad (42)$$

is an independent family. Moreover, $|\sup(\widehat{U_0})| = |\sup(\widehat{U})| - 2D$ so putting $\mathcal{X}_0 = \sup(\widehat{U_0}) \cup \mathcal{U}$ we have

$$|\mathcal{X}_0| = mn - \sum_{i=1}^m c(g_i) - (n-1) - 2D + |\mathcal{U}| = mn - (n-1) - 2D.$$

Define $\psi : \mathcal{K} \rightarrow S$ by $\kappa(j) \mapsto \kappa(j)$, and extend ψ arbitrarily to $\mathcal{K} \cup \mathcal{X}_0$. Let $\mu \in S$ be the value of U_0 determined by ψ . According to Theorem 1.9, there exist $\xi_1, \eta_1, \dots, \xi_D, \eta_D \in S$ such that

$$T_{a_1, b_1}(\xi_1, \eta_1) \cdot \dots \cdot T_{a_D, b_D}(\xi_D, \eta_D) = \kappa(1)\mu^{-1}.$$

Since (42) is an independent family, we can extend ψ to a Γ -equivariant homomorphism $F \rightarrow S$ sending x_i to ξ_i and y_i to η_i for each i , and then $\psi(U) = \kappa(1)$.

Each such mapping ψ thus gives rise to a solution of the original equation (37). Distinct mappings give distinct solutions, because the values of all the variables $x_i(s)$ are determined by the values of the $u_i(s)$ via (40). The number of solutions is therefore at least equal to the number of possible maps ψ , which is at least

$$|S|^{|\mathcal{X}_0|} \geq |S|^{mn-n-2D+1} = \frac{|N|^m}{|S|^{n+2D-1}}.$$

Now $|S| < |S/Z(S)|^2$ ([GLS], §6.1) so

$$|S|^{n+2D-1} < |S|^{2nD} < |S/Z(S)|^{4nD} = |N/Z|^{4D},$$

and the proposition follows.

10 Equations in semisimple groups, 2: powers

Fix a positive integer q . The constants $D, C = C(q)$ and $M = M(q)$ are those appearing in Theorems 1.9 and 1.10, and we put

$$\overline{D} = 4 + 2D, \quad z(q) = M\overline{D}(q + \overline{D}).$$

In this section we establish

Proposition 10.1 *Let N be a quasi-semisimple normal subgroup of a group G and $h_1, \dots, h_m \in G$. Assume that $m \geq z(q)$ and that $|T| > C$ for each non-abelian composition factor T of N . Then the mapping $\psi : N^{(m)} \rightarrow N$ given by*

$$\prod_{i=1}^m (a_i h_i)^q = \psi(a_1, \dots, a_m) \cdot \prod_{i=1}^m h_i^q.$$

is surjective.

Let $H = \langle h_1, \dots, h_m \rangle$. It is clear that ψ depends only on the action of the h_i on N . The action of H on N lifts to an action on the universal cover \tilde{N} , and it will suffice to prove the result for the case where $N = \tilde{N}$. Thus we shall assume that $N = S_1 \times \dots \times S_r$ where each S_i is a quasisimple group; the action of H then permutes the S_i . If $N = N_1 \times \dots \times N_t$ where each N_i is H -invariant, then it is easy to see that $\psi = \psi|_{N_1} \times \dots \times \psi|_{N_t}$; so we may assume in addition that this permutation action is transitive. It follows that $S_i \cong S$ for each i , where S is quasisimple with $|S/Z(S)| > C$.

The explicit expression for ψ is thoroughly unpleasant. Instead of confronting it directly we proceed as follows. For $\mathbf{x}, \mathbf{b} \in N^{(m)}$ and $1 \leq i \leq m$ put

$$a_i(\mathbf{x}, \mathbf{b}) = x_i^{b_i} [b_i, h_i^{-1}],$$

so $a_i(\mathbf{x}, \mathbf{b})h_i = (x_i h_i)^{b_i}$. Then

$$\begin{aligned} \psi(a_1(\mathbf{x}, \mathbf{b}), \dots, a_m(\mathbf{x}, \mathbf{b})) &= \prod_{i=1}^m ((x_i h_i)^q)^{b_i} \cdot \left(\prod_{i=1}^m h_i^q \right)^{-1} \\ &= \prod_{i=1}^m ((x_i h_i)^q)^{b_i} \cdot \left(\prod_{i=1}^m (x_i h_i)^q \right)^{-1} \cdot \psi(\mathbf{x}) \\ &= \prod_{i=1}^m [b_i, (x_i h_i)^{-q}]^{\tau_i(\mathbf{x}\mathbf{h})} \cdot \psi(\mathbf{x}) \end{aligned} \quad (43)$$

where

$$\begin{aligned} \tau_i(\mathbf{x}\mathbf{h}) &= (x_{i-1} h_{i-1})^{-q} \dots (x_1 h_1)^{-q} \\ &= \xi_i(x_1, \dots, x_{i-1}) \tau_i(\mathbf{h}), \end{aligned}$$

say. We shall prove

Proposition 10.2 *Let $N = S^{(r)}$ where S is a quasisimple group with $|S/Z(S)| > C$, and let $H = \langle k_1, \dots, k_m \rangle \leq \text{Aut}(N)$ act transitively on the set of simple factors of N . Suppose that $m \geq z(q)$. Then there exist $y_1, \dots, y_m \in N$ such that*

$$N = \prod_{j=1}^m [N, (y_j k_j)^q].$$

(Here and later, we do not distinguish between an element of N and the inner automorphism it induces).

This suffices to complete the proof of Proposition 10.1. Indeed, suppose we want to solve the equation $\psi(\mathbf{a}) = \kappa$. In view of (43), it will suffice to find \mathbf{x} and \mathbf{b} such that

$$\prod_{i=1}^m [b_i, (x_i h_i)^{-q}]^{\tau_i(\mathbf{x}\mathbf{h})} = \kappa \psi(\mathbf{x})^{-1}. \quad (44)$$

For each i let $k_i \in \text{Aut}(N)$ be induced by $h_i^{-\tau_i(\mathbf{h})}$; it is easy to see that then $\langle k_1, \dots, k_m \rangle$ is the group of automorphisms induced by $\langle h_1, \dots, h_m \rangle$, hence acts

transitively on the S_j . Let $y_1, \dots, y_m \in N$ be as specified in the proposition, and define x_1, \dots, x_m recursively by

$$x_i^{h_i} = [h_i, \xi_i^{-1}] \cdot y_i^{-\tau_i^{-1} \xi_i^{-1}}$$

where $\xi_1 = \tau_1 = 1$ and $\xi_i = \xi_i(x_1, \dots, x_{i-1})$, $\tau_i = \tau_i(\mathbf{h})$ for $i > 1$. According to the proposition, there exists $\mathbf{a} \in N^{(m)}$ such that $\prod_{i=1}^m [a_i, (y_i k_i)^q] = \kappa \psi(\mathbf{x})^{-1}$. Since $y_i k_i$ acts as $(x_i h_i)^{-\tau_i(\mathbf{xh})}$ on N , we may now solve (44) by setting $b_i = a_i^{\tau_i(\mathbf{xh})^{-1}}$.

The rest of this section is devoted to the proof of Proposition 10.2. As in §9, we write $N = S_1 \times \dots \times S_r$ and fix an identification of each S_i with S . For $h \in H$ and $x = (x(1), \dots, x(r)) \in N$ we write

$$x^h = \left(x^{(h_1)^{h(1)}}, \dots, x^{(h_r)^{h(r)}} \right),$$

where $i \mapsto h_i$ is the permutation $\sigma(h^{-1})$ of $\{1, \dots, r\}$ induced by the action of h^{-1} on $\{S_1, \dots, S_r\}$ and $h(i) \in \text{Aut}(S)$ is induced by $h|_{S(h_i)} : S(h_i) \rightarrow S_i$. For $\Delta \subseteq \{1, \dots, r\}$ the projection $N \rightarrow \prod_{i \in \Delta} N_i$ is denoted π_Δ .

The set of fixed points of $\sigma(h)$ is denoted $\text{fix}(h)$, and we write

$$\begin{aligned} \text{fix}^*(i) &= \{j \in \{1, \dots, m\} \mid i \in \text{fix}(k_j^q)\}, \\ \lambda(\Delta) &= m |\Delta| - \sum_{i \in \Delta} |\text{fix}^*(i)|. \end{aligned}$$

Thus $\lambda(\Delta)$ is the number of pairs (i, j) with $i \in \Delta$ such that k_j^q moves S_i .

Put $G_1 = \langle k_1^q, \dots, k_m^q \rangle$ and let Ω be an orbit of $\sigma(G_1)$ on $\{1, \dots, r\}$. We say that Ω is of *type I* if $\lambda(\Omega) < \overline{D} |\Omega|$, of *type II* otherwise.

When Ω is of type I there exists at least one $i \in \Omega$ for which $|\text{fix}^*(i)| > m - \overline{D}$; we choose such a value of i and denote it i_Ω . Put

$$\mathcal{S} = \bigcup \{(\Omega, j) \mid j \in \text{fix}^*(i_\Omega)\}$$

where Ω ranges over all the G_1 -orbits of type I. Two pairs (Ω, j) and (Ω', j') will be called *independent* if either $j \neq j'$ or $j = j'$ and i_Ω and $i_{\Omega'}$ lie in distinct orbits of k_j ; a subset of \mathcal{S} is independent if its members are pairwise independent.

Lemma 10.3 *Suppose that $m \geq z(q)$. Then for each G_1 -orbit Ω of type I there exist an interval $J_\Omega \subseteq \text{fix}^*(i_\Omega)$ and a subset $I_\Omega \subseteq J_\Omega$ such that*

- (i) $|I_\Omega| = M$,
- (ii) *the set*

$$\mathcal{T} = \{(\Omega, j) \in \mathcal{S} \mid j \in I_\Omega\}$$

is independent.

This lemma will be proved below. Now let Ω be a G_1 -orbit of type I. If $j \in I_\Omega$ then $\sigma(k_j)^q$ fixes i_Ω , so the k_j -cycle $C(\Omega, j)$ of i_Ω has length $e(\Omega, j)$, say, dividing q . Put $q_{\Omega j} = q/e(\Omega, j)$ and let $\beta_{\Omega j}$ denote the automorphism of S_{i_Ω} induced by the action of $k_j^{e(\Omega, j)}$. According to Theorem 1.10, we may choose elements $x_{\Omega j} \in S_{i_\Omega}$ so that

$$S_{i_\Omega} = \prod_{j \in I_\Omega} [S_{i_\Omega}, (x_{\Omega j} \beta_{\Omega j})^{q_{\Omega j}}]. \quad (45)$$

In this way we obtain a family of elements $x_{\Omega j} \in S_{i_\Omega}$ as (Ω, j) ranges over \mathcal{T} . Now for each $j \in \{1, \dots, m\}$ let

$$y_j = \prod_{I_\Omega \ni j} x_{\Omega j} \in \prod_{I_\Omega \ni j} S_{i_\Omega}$$

($y_j = 1$ if the range of the product is empty). The independence of \mathcal{T} ensures that if $j \in I_\Omega$ then $\pi_{C(\Omega, j)}(y_j) = x_{\Omega j}$, and hence that $(y_j k_j)^{e(\Omega, j)}$ acts on S_{i_Ω} as $x_{\Omega j} \beta_{\Omega j}$. Thus writing $g_j = (y_j k_j)^q$ for each j , we have

$$S_{i_\Omega} = \prod_{j \in I_\Omega} [S_{i_\Omega}, g_j] = \prod_{j \in J_\Omega} [S_{i_\Omega}, g_j] \quad (46)$$

for each G_1 -orbit Ω of type I.

Now put $G = \langle g_1, \dots, g_m \rangle$, and note that $\sigma(g_j) = \sigma(k_j^q)$ for each j , $\sigma(G) = \sigma(G_1)$. For each G -orbit Ω let $N_\Omega = \prod_{i \in \Omega} S_i$. Then N is the direct product of the N_Ω , each of which is invariant under G . Thus to prove Proposition 10.2 it will suffice to show that

$$N_\Omega = \prod_{j=1}^m [N_\Omega, g_j] \quad (47)$$

for each G -orbit Ω .

Case 1: where Ω is of type II. Assume for ease of notation that $\Omega = \{1, \dots, n\}$. Then

$$\lambda(\Omega) = mn - \sum_{i=1}^n |\text{fix}^*(i)| \geq n\overline{D},$$

where $\text{fix}^*(i) = \{j \mid g_j i = i\}$. Note that this entails $n \geq 2$. Let $c(g_j)$ denote the number of cycles of $\sigma(g_j)$. Then

$$\begin{aligned} \sum_{i=1}^m c(g_j) &\leq \frac{1}{2} \sum_{j=1}^m (n + |\text{fix}_\Omega(g_j)|) \\ &= \frac{1}{2} (mn + \sum_{i=1}^n |\text{fix}^*(i)|) \\ &= mn - \frac{1}{2} \lambda(\Omega) \leq mn - \frac{1}{2} \overline{D}n = (m-2)n - nD. \end{aligned}$$

Since $n \geq 2$, the identity (47) now follows from Proposition 9.1.

Case 2: where Ω is of type I. Say $\Omega = \{1, \dots, n\}$, and that $i_\Omega = 1$. Suppose that the interval J_Ω is $\{l, l+1, \dots, p\}$, so $\sigma(g_j)$ fixes 1 for $l \leq j \leq p$.

Given $\kappa \in N_\Omega$ we have to solve the equation

$$\kappa = x_1 \dots x_m \quad (48)$$

subject to the conditions

$$x_i \in [N_\Omega, g_i], \quad (49)$$

$i = 1, \dots, m$. Let Ω_i denote the set of orbits of $\sigma(g_i)$ in Ω ; for $\Delta \in \Omega_i$ write k_Δ for its first member and put $n(\Delta) = |\Delta|$. As shown in the proof of Proposition 9.1, the condition (49) is satisfied if and only if

$$\prod_{j=0}^{n(\Delta)-1} x_i(g_i^j k_\Delta)^{g_i^j(k_\Delta)} \in [S, \beta_i(\Delta)]$$

for each $\Delta \in \Omega_i$, where $\beta_i(\Delta) = g_i^{n(\Delta)}(k_\Delta)$ is the automorphism of S induced by the action of $g_i^{n(\Delta)}$ on $S_{k(\Delta)}$ (see formula $(H_{i,\Delta})$ in §9).

Write (48) as the system of equations \mathcal{F} :

$$\kappa(s) = x_1(s) \dots x_m(s), \quad (F_s)$$

$s = 1, \dots, n$. For each s , let F'_s be the equation obtained from F_s as follows: for each pair (i, Δ) with $\Delta \in \Omega_i$ and $k_\Delta = s$, replace $x_i(s)$ by the expression

$$V_i(\Delta) \cdot \left(\prod_{j=1}^{n(\Delta)-1} x_i(g_i^j k_\Delta)^{g_i^j(k_\Delta)} \right)^{-1}, \quad (50)$$

where $V_i(\Delta)$ is a new symbol. Note that for $i \in J_\Omega$ we have an orbit $\Delta = \{1\} \in \Omega_i$, so the first equation becomes

$$\kappa(1) = \bar{x}_1(1) \dots \bar{x}_{l-1}(1) \cdot \prod_{i=l}^p V_i(\{1\}) \cdot \bar{x}_{p+1}(1) \dots \bar{x}_m(1), \quad (F'_1)$$

where $\bar{x}_i(1)$ stands for the expression (50) with Δ the g_i -orbit of 1.

The resulting system \mathcal{F}' of equations contains the unknowns $V_i(\Delta)$ for $\Delta \in \Omega_i$ and the $x_i(s)$ for every s not of the form k_Δ , $\Delta \in \Omega_i$; each such $x_i(s)$ now occurs exactly once with its inverse. We are required to solve \mathcal{F}' with each $x_i(s) \in S$ and each $V_i(\Delta) \in [S, \beta_i(\Delta)]$.

Next, we reduce \mathcal{F}' to a single equation using the procedure described in the proof of Proposition 9.1. That is, if a term $x = x_i(j)^{\pm 1}$ appears in F'_1 but its inverse does not, then x^{-1} appears in some F'_l , $l \neq 1$. Solve F'_l for x and substitute the resulting expression in F'_1 ; cross out the equation F'_l , and iterate.

As we saw in the preceding section, the transitivity of $\sigma(G)$ ensures that after $n - 1$ such steps the equations F'_2, \dots, F'_n will have been eliminated.

Since the ‘middle part’ of F'_1 is unaffected by this process, the resulting equation takes the form

$$\kappa(1) = A \cdot \prod_{i=l}^p V_i(\{1\}) \cdot B$$

where $A \cdot B$ is the product, in some order, of certain terms $x_i(j)^\varepsilon$, all the $V_i(\Delta)$ with $\Delta \neq \{1\}$ when $l \leq i \leq p$, and $\kappa(2)^{-1}, \dots, \kappa(n)^{-1}$, possibly with an automorphism attached. Setting each such $x_i(j)$ and each such $V_i(\Delta)$ equal to 1, we are reduced to solving

$$\prod_{i=l}^p V_i(\{1\}) = \kappa^*$$

for a certain $\kappa^* \in S$, subject to the conditions $V_i(\{1\}) \in [S, \beta_i(\{1\})]$ for $l \leq i \leq p$. But $\beta_i(\{1\})$ is just the automorphism induced by the action of g_i on S_1 ; the solubility of this equation is therefore assured by (46).

This completes the proof.

It remains to give the

Proof of Lemma 10.3 Let \mathcal{O} denote the set of all G_1 -orbits of type I. For each $\Omega \in \mathcal{O}$ we are given $i_\Omega \in \Omega$ such that $\sigma(k_j)^q$ fixes i_Ω for all but at most $\overline{D} - 1$ values of j ; thus the set $\text{fix}^*(i_\Omega)$ is the union of at most \overline{D} intervals.

We make the following

Claim: Let $X \subseteq \{1, \dots, m\}$ be a subset with $|X| \geq q + \overline{D}$. Then there exists a mapping $j : \mathcal{O} \rightarrow X$ such that

$$\mathcal{S}_X = \{(\Omega, j(\Omega)) \mid \Omega \in \mathcal{O}\}$$

is an independent subset of \mathcal{S} .

Accepting the claim for now, partition the sequence $\{1, \dots, m\}$ into $M\overline{D}$ intervals $X(1), \dots, X(M\overline{D})$ of length at least $q + \overline{D}$, and put

$$\tilde{\mathcal{T}} = \bigcup_{i=1}^{M\overline{D}} \mathcal{S}_{X(i)}.$$

This is evidently an independent set. Now fix $\Omega \in \mathcal{O}$ and consider the set $\tilde{\mathcal{T}}_\Omega = \{j \mid (\Omega, j) \in \tilde{\mathcal{T}}\}$. This meets each $X(i)$, so has cardinality at least $M\overline{D}$. Since $\tilde{\mathcal{T}}_\Omega \subseteq \text{fix}^*(i_\Omega)$ it follows that $|\tilde{\mathcal{T}}_\Omega \cap J| \geq M$ for at least one of the (at most) \overline{D} intervals J that make up $\text{fix}^*(i_\Omega)$. Put $J_\Omega = J$ and let I_Ω be any subset of $\tilde{\mathcal{T}}_\Omega \cap J$ of size M . These then satisfy all the requirements of the lemma.

To prove the *Claim*, we will apply Hall’s ‘marriage theorem’ (see e.g. [PB], Chapter 22). The ‘men’ are pairs (Δ, j) where $j \in X$ and Δ is an orbit of $\sigma(k_j)$

with $|\Delta| \mid q$. The set of ‘women’ is just \mathcal{O} , and we say that Ω ‘knows’ (Δ, j) (and vice versa) precisely when $i_\Omega \in \Delta$. Evidently each man knows at most q women; while each woman Ω knows at least q men, namely the

$$(C(\Omega, j), j), \quad j \in X \cap \text{fix}^*(i_\Omega)$$

where $C(\Omega, j)$ is the $\sigma(k_j)$ -cycle containing i_Ω . It follows (counting possible ‘couples’ in two ways) that for every n , each set of n women collectively knows at least n men. Hall’s theorem now ensures that each woman Ω can find a husband $(\Delta(\Omega), j(\Omega))$ with $i_\Omega \in \Delta(\Omega)$. The monogamy rule means that if $\Omega \neq \Omega'$ then $(\Delta(\Omega), j(\Omega)) \neq (\Delta(\Omega'), j(\Omega'))$; this is precisely the statement that the pairs $(\Omega, j(\Omega))$ and $(\Omega', j(\Omega'))$ are independent.

11 Equations in semisimple groups, 3: twisted commutators

Theorem 1.9, stated in the Introduction, asserts that every element of any finite quasisimple group can be written as a product of boundedly many twisted commutators. Here we generalize this result. Recall the notation

$$T_{\alpha, \beta}(x, y) = x^{-1} y^{-1} x^\alpha y^\beta,$$

and let D be the absolute constant given in Theorem 1.9.

Proposition 11.1 *Let N be a quasi-semisimple group and α_l, β_l ($l = 1, 2, \dots, D$) arbitrary automorphisms of N . Then*

$$\prod_{i=1}^D T_{\alpha_i, \beta_i}(N, N) = N. \quad (51)$$

The universal cover of N is a direct product $\tilde{N} = S_1 \times \dots \times S_n$ where each S_i is a quasisimple group of universal type. Each automorphism of N lifts to one of \tilde{N} , so it will suffice to prove the result in the case $N = \tilde{N}$, which we assume henceforth.

Let $G = \langle \alpha_l, \beta_l \mid 1 \leq l \leq D \rangle$ be the subgroup of $\text{Aut}(N)$ generated by the given automorphisms. Then G permutes the factors S_1, \dots, S_n , with orbits Λ_i say. Now N is the direct product of the subgroups $N(i) = \prod_{j \in \Lambda_i} S_j$, on each of which G acts by restriction, and it will suffice to prove (51) with $N(i)$ in place of N , for each i . Thus we may, and shall, assume that the permutation action of G on $\{S_1, \dots, S_n\}$ is transitive. As in the preceding sections, we shall write

$$S_i^{\alpha^{-1}} = S_{\alpha i} \quad (\alpha \in G, 1 \leq i \leq n).$$

Since this action of G is transitive, the groups S_i are all isomorphic; we fix an identification of each S_i with a fixed quasisimple group S . Thus elements of N will be written in the form

$$x = (x(1), x(2), \dots, x(n))$$

with each $x(i) \in S$, and the action of G takes the form

$$x^\alpha = (x^{(\alpha 1)^{\alpha(1)}}, x^{(\alpha 2)^{\alpha(2)}}, \dots, x^{(\alpha n)^{\alpha(n)}});$$

here $\alpha(1), \dots, \alpha(n) \in \text{Aut}(S)$ depend on $\alpha \in G$ (and the fixed identifications $S_i \rightarrow S$). We put

$$\Gamma = \langle \alpha_i(s), \beta_i(s) \mid 1 \leq i \leq D, 1 \leq s \leq n \rangle \leq \text{Aut}(S).$$

Let $\kappa \in N$. We have to show that there exist $\kappa_1, \dots, \kappa_D \in N$ such that

$$\kappa = \kappa_1 \dots \kappa_D \quad (52)$$

and such that for each i there exist $x_i, y_i \in N$ with

$$\kappa_i = x_i^{-1} y_i^{-1} x_i^{\alpha_i} y_i^{\beta_i}. \quad (53)$$

To begin with, we fix i and analyse the equation (53). This equation is solvable in N if and only if there exist elements $x_i(s), y_i(s) \in S$ ($s = 1, \dots, n$) such that (writing $\alpha = \alpha_i, \beta = \beta_i$)

$$\kappa_i(s) = x_i(s)^{-1} y_i(s)^{-1} x_i^{\alpha(s)} y_i^{\beta(s)} \quad (E_s)$$

holds for $s = 1, \dots, n$. We consider $\mathcal{E} = (E_1, \dots, E_n)$ as a system of simultaneous equations in the unknowns $x_i(s), y_i(s)$.

Put

$$G_i = \langle \alpha_i, \beta_i \rangle$$

and let $\Omega = \Omega_i$ denote the set of orbits of G_i on $\{1, \dots, n\}$. The system \mathcal{E} breaks up into $|\Omega|$ independent systems of equations, one for each orbit $\Delta \in \Omega$:

$$\mathcal{E}_\Delta = (E_s)_{s \in \Delta}.$$

We fix an orbit Δ of size n_Δ , and introduce the alphabet $X \cup P$ where

$$\begin{aligned} X &= X_\Delta = \{x_i(s), y_i(s) \mid s \in \Delta\} \\ P &= P_\Delta = \{\kappa_i(s) \mid s \in \Delta\}; \end{aligned}$$

for now the elements of $X \cup P$ are considered as abstract symbols. Here P is the set of *parameters* and X is the set of *variables*. Let $F_\Delta = F_\Gamma(X_\Delta \cup P_\Delta)$ be the free Γ -group, defined in Section 8. For $x \in X^{\pm 1} \cup P^{\pm 1}$ we will write x^* to denote an arbitrary element of the form x^γ , $\gamma \in \Gamma$.

We consider the right-hand sides of the equations E_s in \mathcal{E}_Δ as words on the alphabet $X^\Gamma \cup P^\Gamma$. Note that each variable occurs exactly once, as does its inverse, in the system \mathcal{E}_Δ .

Let $k = k_\Delta$ be the first symbol in Δ . We shall modify \mathcal{E}_Δ by the familiar process of eliminating variables. Suppose that x^* occurs in the first equation E_k , where $x \in X \cup X^{-1}$; then a term x^{-*} occurs in some equation E_l . If $l \neq k$, solve E_l for x and substitute the resulting value of x in E_k . Let us call this process

a *substitution* ($l \rightarrow k$). Each substitution reduces by one both the number of variables and the number of equations in the system \mathcal{E}_Δ .

We claim that it is possible to apply $n_\Delta - 1$ substitutions and thus reach an equivalent system consisting of the single equation

$$\kappa_i(k_\Delta) = U_\Delta, \quad (54)$$

where U_Δ is a certain word on $X^\Gamma \cup P^\Gamma$. This follows just as in §9 from the fact that G_i acts transitively on Δ .

As in §8, let $\mathcal{M} = \mathcal{M}_\Delta$ denote the free monoid on $X^{\pm\Gamma} \cup P^{\pm\Gamma}$. Recall that for words $U, U' \in \mathcal{M}$ the expression $U =_F U'$ means that U and U' take the same value \overline{U} in the group F , and that \widehat{U} denotes the word in W , the free monoid on $X \cup X^{-1}$, obtained from U when all symbols from $P^{\pm\Gamma}$ are deleted and x^γ is replaced by x for each $x \in X \cup X^{-1}$, $\gamma \in \Gamma$.

Lemma 11.2 *There exist $x = x_\Delta$, $y = y_\Delta$, $V = V_\Delta \in \mathcal{M}$ and $a = a_\Delta$, $b = b_\Delta \in \Gamma$ such that*

$$U_\Delta =_F T_{a,b}(x, y) \cdot V,$$

the family $\{\overline{x}, \overline{y}\} \cup \sup(\widehat{V}) \cup P$ is independent, each $\kappa_i(s)$ for $s \in \Delta \setminus \{k_\Delta\}$ occurs exactly once in V with exponent $-\gamma$ for some $\gamma \in \Gamma$, and $\kappa_i(k_\Delta)$ does not occur in V .

Proof. Let U_1 be the word on the right-hand side of E_k , and let U_f denote the word obtained from U_1 after $f - 1$ substitutions have been carried out. A substitution ($l \rightarrow k$) has one of the following effects (we drop the subscript i from $x(s)$, $y(s)$, α and β):

$\frac{l}{s}$	$\frac{U_f}{A \cdot x(s)^* \cdot B}$	$\frac{U_{f+1}}{A \cdot y(l)^{-*} x^{(\alpha l)^*} y^{(\beta l)^*} \kappa_i(l)^{-*} \cdot B}$
$\alpha^{-1} s$	$A \cdot x(s)^{-*} \cdot B$	$A \cdot y^{(\beta l)^*} \kappa_i(l)^{-*} x(l)^{-*} y(l)^{-*} \cdot B$
s	$A \cdot y(s)^* \cdot B$	$A \cdot x^{(\alpha l)^*} y^{(\beta l)^*} \kappa_i(l)^{-*} x(l)^{-*} \cdot B$
$\beta^{-1} s$	$A \cdot y(s)^{-*} \cdot B$	$A \cdot \kappa_i(l)^{-*} x(l)^{-*} y(l)^{-*} x^{(\alpha l)^*} \cdot B$

Since each variable occurs exactly once with its inverse in the system \mathcal{E}_Δ , it is easy to see that the same holds for the final word $U_\Delta = U_{n_\Delta}$, except for the $n_\Delta - 1$ matching pairs x, x^{-1} that have been eliminated. Thus the word \widehat{U}_Δ is balanced.

We claim also that $\widehat{U}_\Delta \neq_F 1$. To see this, let $\Phi = \langle \xi, \eta \rangle$ be a the free nilpotent group of class two on two free generators, and define a (monoid) homomorphism $\theta : W \rightarrow \Phi$ by

$$x(s)^\varepsilon \mapsto \xi^\varepsilon, \quad y(s)^\varepsilon \mapsto \eta^\varepsilon$$

($\varepsilon = \pm 1$, $s \in \Delta$). Now we can write $\theta(\widehat{U_f}) = \theta(\widehat{A})\theta(z)\theta(\widehat{B})$ where z is one of $x(s)^{\pm 1}$, $y(s)^{\pm 1}$, and we see that in the four cases listed we get, respectively,

$$\begin{aligned}\theta(\widehat{U_{f+1}}) &= \theta(\widehat{A})\theta(z) \cdot [\xi, \eta] \cdot \theta(\widehat{B}) \\ \theta(\widehat{U_{f+1}}) &= \theta(\widehat{A})\theta(z) \cdot [\xi^{-1}, \eta^{-1}] \cdot \theta(\widehat{B}) \\ \theta(\widehat{U_{f+1}}) &= \theta(\widehat{A})\theta(z) \cdot [\eta, \xi^{-1}] \cdot \theta(\widehat{B}) \\ \theta(\widehat{U_{f+1}}) &= \theta(\widehat{A})\theta(z) \cdot [\eta^{-1}, \xi] \cdot \theta(\widehat{B})\end{aligned}$$

each of which is equal to $[\xi, \eta]\theta(\widehat{U_f})$. As $\theta(\widehat{U_1}) = [\xi, \eta]$ it follows that $\theta(\widehat{U_\Delta}) = [\xi, \eta]^{n_\Delta} \neq 1$. Since θ factors through F this establishes the claim.

Thus U_Δ satisfies the conditions of Proposition 8.2. This now gives the result, provided only that the multiplicities of the $\kappa_i(s)$ in U_Δ are as described in the statement. But this is clear, since each substitution ($l \rightarrow k$) as above introduces the term $\kappa_i(l)^{-*}$ (and no other terms from $P \cup P^{-1}$), and the label l runs over the set $\Delta \setminus \{k_\Delta\}$ as f goes from 1 to $n_\Delta - 1$. ■

The preceding reduction now shows that the equation (53) is solvable in N if and only if for each orbit $\Delta \in \Omega_i$ there exists a Γ -homomorphism

$$\phi_\Delta : F_\Delta \rightarrow S$$

sending each symbol $\kappa_i(s)$ to the element with the same name in S and satisfying

$$\phi_\Delta(T_{a_\Delta, b_\Delta}(x_\Delta, y_\Delta)V_\Delta) = \kappa_i(k_\Delta).$$

Now put $Z_\Delta = \sup(\widehat{V_\Delta})$, and consider a new alphabet $Y \cup P \cup \mathcal{K}$ where

$$\begin{aligned}Y &= \bigcup_{i=1}^D \bigcup_{\Delta \in \Omega_i} Z_\Delta \cup \{x_\Delta, y_\Delta\} \\ P &= \bigcup_{i=1}^D \bigcup_{\Delta \in \Omega_i} P_\Delta = \{\kappa_i(s) \mid 1 \leq i \leq D, 1 \leq s \leq n\} \\ \mathcal{K} &= \{\kappa(1), \dots, \kappa(n)\}.\end{aligned}$$

The equation (52) is equivalent to the system of equations $\mathcal{F} = (F_1, \dots, F_n)$:

$$\kappa(s) = \kappa_1(s) \dots \kappa_D(s). \quad (F_s)$$

For each pair (i, Δ) with $\Delta \in \Omega_i$ we substitute the expression $T_{a_\Delta, b_\Delta}(x_\Delta, y_\Delta)V_\Delta$ for $\kappa_i(k_\Delta)$ in the equation F_{k_Δ} , to obtain a system $\mathcal{F}' = (F'_1, \dots, F'_n)$:

$$\kappa(s) = W_s \quad (F'_s)$$

where W_s is a certain word on the alphabet $Y^\Gamma \cup P^\Gamma$. Now recall that V_Δ contains $\kappa_i(s)^{-*}$ exactly once for each $s \in \Delta \setminus \{k_\Delta\}$, and no other terms from

$P^{\pm\Gamma}$; it follows that $W_1W_2\ldots W_n$ contains the terms $\kappa_i(s), \kappa_i(s)^{-*}$ once each whenever $s \notin \{k_\Delta \mid \Delta \in \Omega_i\}$, and no other terms from $P^{\pm\Gamma}$.

We now repeat the elimination procedure used above. Suppose that $\mu \in P \cup P^{-1}$ and μ^* occurs in F'_1 while μ^{-*} occurs in F'_l for some $l \neq 1$. Solve F'_l for μ and substitute the resulting expression into F'_1 . It is easy to see that two equations F'_p, F'_q are 'linked', in the sense that they share a parameter from P , if and only if there exists i such that p and q lie in the same orbit of G_i . Since the G_i generate G which is transitive on $\{1, \dots, n\}$, it follows as before that we can perform $n-1$ such substitutions and obtain an equivalent system consisting of one equation

$$\kappa(1) = V. \quad (55)$$

Each substitution ($l \rightarrow 1$) eliminates a pair $\kappa_i(l), \kappa_i(l)^{-1}$ and introduces into the right-hand member of F'_1 both a term $\kappa(l)^{-1}$ and all the terms $T_\Delta = T_{a_\Delta, b_\Delta}(x_\Delta, y_\Delta)$ that appear in F'_l (ignoring exponents from Γ). It follows that V contains each of the terms $\kappa(2)^{-1}, \dots, \kappa(n)^{-1}$ exactly once, and each of the terms T_Δ ($\Delta \in \Omega_i, 1 \leq i \leq D$) exactly once. The other factors of V (still ignoring exponents from Γ) all belong to $P^{\pm 1} \cup \bigcup_\Delta Z_\Delta^{\pm 1}$.

Let $\mathcal{X} = \{x_i(s) \mid 1 \leq i \leq D, 1 \leq s \leq n\}$. Recall now (Lemma 11.2) that each of the families $\{\overline{x_\Delta}, \overline{y_\Delta}\} \cup Z_\Delta \cup P_\Delta$ is independent in the free Γ -group F_Δ . This implies that the family

$$\bigcup_{\Delta \in \Omega_i, 1 \leq i \leq D} (\{\overline{x_\Delta}, \overline{y_\Delta}\} \cup Z_\Delta) \cup P \cup \mathcal{K}$$

is independent in the free Γ -group F on $\mathcal{X} \cup P \cup \mathcal{K}$. Hence for any choice of elements $\xi_\Delta, \eta_\Delta \in S$ there is a Γ -equivariant homomorphism $\phi_{\xi, \eta} : F \rightarrow S$ sending x_Δ to ξ_Δ , y_Δ to η_Δ , each symbol $\kappa(i)$ to the given element $\kappa(i)$ of S , and each term of $P \cup \bigcup_\Delta Z_\Delta$ that appears in V to 1. Then

$$\phi_{\xi, \eta}(\kappa(1)) = \kappa(1),$$

while

$$\phi_{\xi, \eta}(V) = h_0 \prod_{\Delta \in \Omega_i, 1 \leq i \leq D} T_{a_\Delta, b_\Delta}(\xi_\Delta, \eta_\Delta)^{\gamma_\Delta h_\Delta}$$

(in some order) where the $\gamma_\Delta \in \Gamma$ and $h_0, h_\Delta \in S$ do not depend on ξ, η .

Using the identity

$$T_{a, b}(x, y)^\gamma = T_{a^\gamma, b^\gamma}(x^\gamma, y^\gamma)$$

we rewrite the above as

$$h_1^{-1} \phi_{\xi, \eta}(V) = \prod_{\Delta \in \Omega_i, 1 \leq i \leq D} T_{a'_\Delta, b'_\Delta}(\xi'_\Delta, \eta'_\Delta) \quad (56)$$

where $a'_\Delta, b'_\Delta \in \text{Aut}(S)$ and $\xi'_\Delta, \eta'_\Delta$ are the images of ξ_Δ, η_Δ under certain fixed automorphisms of S .

Now Theorem 1.9 asserts that $S = \prod T_{a'_\Delta, b'_\Delta}(S, S)$ provided there are at least D factors in the product. Hence we can choose values for ξ_Δ, η_Δ in S so that the product on the right of (56) takes the value $h_1^{-1}\kappa(1)$.

The original equation (52) is now solved subject to the conditions (53) by giving each unknown $x_i(s)$ the value $\phi_{\xi, \eta}(x_i(s))$. This completes the proof of Proposition 11.1.

References

- [PB] M. Aigner and G. M. Ziegler, *Proofs from THE BOOK*, 2nd ed., Springer, Berlin, 2001.
- [BCP] L. Babai, P. J. Cameron and P. P. Pálffy, On the orders of primitive groups with restricted nonabelian composition factors, *J. Algebra* **79** (1982), 161-168
- [Cm] P. J. Cameron. *Permutation groups*. LMS Student Texts **45**, Cambridge Univ. Press, Cambridge, 1999.
- [GSS] D. Gluck, A. Seress and A. Shalev, Bases for primitive permutation groups and a conjecture of Babai. *J. Algebra* **199** (1998), 367-378.
- [G] D. Gorenstein, *Finite simple groups*, Plenum Press, New York, 1982.
- [GLS] D. Gorenstein, R. Lyons and R. Solomon, *The classification of the finite simple groups 3*, AMS Mathematical surveys and monographs **40**, 1998.
- [Hm] Y.O. Hamidoune, An application of connectivity theory in graphs to factorization of elements in groups. *European J. Combin.* **2** (1981), 349-355.
- [Ho] Enumerating perfect groups, *J. London Math. Soc.* (2) **39** (1989), 67-78.
- [Hr] B. Hartley, Subgroups of finite index in profinite groups. *Math. Zeit.* **168** (1979), 71-76.
- [K] *Kourovka Notebook*, 7th ed., Novosibirsk, 1980.
- [LP] M. W. Liebeck and L. Pyber, Finite linear groups and bounded generation, *Duke Math. J.* **107** (2001), 159-171.
- [LS1] M. W. Liebeck and A. Shalev, Simple groups, permutation groups, and probability, *J. Amer. Math. Soc.* **12** (1999), 497-520.
- [LS2] M. W. Liebeck and A. Shalev, Diameters of finite simple groups: sharp bounds and applications, *Annals of Math.* **154** (2001), 383-406.
- [MS] A. Mann and A. Shalev, Simple groups, maximal subgroups and probabilistic aspects of profinite groups, *Israel J. Math.* **96** (1996), 449-468.

- [MZ] C. Martinez, E. Zelmanov, Products of powers in finite simple groups. *Israel J. Math.* **96** (1996), 469-479.
- [N1] N. Nikolov, Power subgroups of profinite groups. *D.Phil. thesis*, University of Oxford, 2002.
- [N2] N. Nikolov, On the commutator width of perfect groups, *Bull. London Math. Soc.* **36** (2004), 30-36.
- [NS] N. Nikolov and D. Segal, On finitely generated profinite groups, II: products in quasisimple groups, *this journal*,
- [RZ] L. Ribes and P. A. Zalesskii, *Profinite groups*. Ergebnisse der Math. **40**, Springer, Berlin – Heidelberg, 2000.
- [R] V. A. Roman'kov, Width of verbal subgroups in solvable groups, *Algebra i Logika* **21** (1982), 60-72 (*Russian*); *Algebra and Logic* **21** (1982), 41-49 (*English*).
- [SW] J. Saxl, J. S. Wilson, A note on powers in simple groups. *Math. Proc. Camb. Phil. Soc.* **122** (1997), 91-94.
- [Sg] D. Segal, Closed subgroups of profinite groups. *Proc. London Math. Soc.* (3) **81** (2000), 29-54.
- [Sr] J-P. Serre, *Galois Cohomology*. Springer Verlag, Berlin-Heidelberg, 1997.
- [W] J. S Wilson, On simple pseudofinite groups. *J. London Math. Soc.* (2) **51** (1995), 471-490.
- [Z] E. I. Zel'manov, 'Solution of the restricted Burnside problem for groups of odd exponent', *Izv. Akad. Nauk. USSR* **54** (1990), 42-59; 'Solution of the restricted Burnside problem for 2-groups', *Mat. Sb.* **182** (1991), 568-592 (*Russian*); *Math. USSR-Sb.* **72** (1992), 543-565 (*English*).

Nikolay Nikolov
 New College
 Oxford OX1 3BN
 UK.

Dan Segal
 All Souls College
 Oxford OX1 4AL
 UK.